PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/82, 15/29, C07K 14/415, C12N 5/10, C07K 16/16, C12Q 1/68, G01N 33/563, A01H 5/00

A2

(11) International Publication Number: **WO 00/06754**

(43) International Publication Date: 10 February 2000 (10.02.00)

(21) International Application Number: PCT/NL99/00491

(22) International Filing Date: 30 July 1999 (30.07.99)

(30) Priority Data:
PCT/NL98/00445 31 July 1998 (31.07.98) NL

(71) Applicants *(for all designated States except US)*: CENTRUM VOOR PLANTENVEREDELINGS-EN REPRODUCTIEONDERZOEK (CPRO-DLO) [NL/NL]; P.O. Box 16, NL-6700 AA Wageningen (NL). LANDBOUWUNIVERSITEIT WAGENINGEN [NL/NL]; P.O. Box 8123, NL-6700 ES Wageningen (NL).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: VAN DER VOSSEN, Edwin, Andries, Gerard [NL/NL]; Bouwstraat 1, NL-3572 SN Utrecht (NL). VAN DER VOORT, Jeroen, Nicolaas, Albert, Maria, Rouppe [NL/NL]; Omval 91, NL-1096 AA Amsterdam (NL). LANKHORST, Rene, Marcel, Klein [NL/NL]; Nijburgsestraat 43, NL-6668 AZ Randwijk (NL). BAKKER, Jaap [NL/NL]; Geertjesweg 122, NL-6704 PD Wageningen (NL). STIEKEMA, Wilhelmus, Johannes [NL/NL]; Leonard Roggeveenstraat 21, NL-6708 SL Wageningen (NL).

(74) Agent: DE BRUIJN, Leendert, C.; Nederlandsch Octrooibureau, Scheveningseweg 82, P.O. Box 29720, NL-2502 LS The Hague (NL).

(81) Designated States: JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: ENGINEERING NEMATODE RESISTANCE IN SOLANACAE

(57) Abstract

The present invention relates to the *Gpa2* resistance gene from potato conferring resistance to phytopathogenic nematodes of the genus *Globodera*. It further relates to methods and materials employing the gene and processes for identifying related genes. Finally the invention relates to polypeptides encoded by said resistance genes and the use of said polypeptides.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

ENGINEERING NEMATODE RESISTANCE IN SOLANACAE

FIELD OF THE INVENTION

The present invention relates to the *Gpa2* resistance gene from potato conferring resistance to phytopathogenic nematodes of the genus *Globodera*. It further relates to methods and materials employing the gene and processes for identifying related genes. Finally the invention relates to polypeptides encoded by said resistance genes and the use of said polypeptides.

BACKGROUND OF THE INVENTION

Plant defense

Most plants are susceptible to infection by pathogens such as nematodes and develop various undesirable disease symptoms upon infection which cause retarded growth, reduced yield and consequently economical loss to farmers. The plants respond to infection with several defense mechanisms including production of phytoalexins, deposition of lignin-like material, accumulation of cell wall hydroxyproline-rich glycoproteins, expression of pathogenesis related proteins (PR-proteins) and an increase in the activity of several lytic enzymes. Some of these responses can be induced not only directly by infection but also in some cases by exposure to exogenous chemicals such as ethylene. The full capacity of the defense mechanism of the plant is, however, normally delayed in relation to the onset of infection, and thus, the plant may be severely injured before its defense mechanism reaches its maximum capacity. Also, the defense mechanism of the plant may not in itself be sufficiently strong to effectively combat the infectious organism. This is in particular true for cultivated plants which have often been cultivated with the aim of increasing the yield, decreasing the climate susceptibility, decreasing the nutrient demand etc. Therefore, a normal and necessary procedure is to treat infected plants or plants susceptible to infection with a chemical either as a prophylactic treatment or shortly after infection. The use of a chemical treatment is neither desirable from an ecological nor from an economic point of view. Another procedure to combat the infectious organism is crop rotation. However, this is not able to fully overcome the problem. It would therefore be desirable to be able to enhance the

defense of the host plant itself by introducing new and/or improved genes by genetic engineering. The advantageous effect of the latter strategy would be the immediate inhibition of a phytopathogenic attack, leading to a retarded epidemic establishment of the infecting organisms in genetically engineered plant crops and thus an overall reduction in the effect of the infection.

One of the phytopathogenic organisms which are most wide spread and which are pathogenic to potato are the potato cyst nematodes (PCN) *Globodera pallida* and *G. rostochiensis*. These nematodes cause considerable losses to potato crop growing, up to 10% of the annual yield world wide. Because cysts are very persistent to chemical treatment and can survive for several years in the soil, the use of nematicides and crop rotation are only moderately effective. The present invention circumvents these drawbacks in the control of PCN.

Durability of PCN resistance

The durability of the resistance is determined by the extent of variation at (a)virulence loci which occur among the pathogen biotypes and the ability of the pathogen to generate novel specificities at (a)virulence loci. For PCN, the variation at (a)virulence loci is for the majority determined by the original founders which have been introduced into Europe. PCN are endemic in the Andes region of South-America where they coevolved with their Solanaceous hosts. They are thought to have been introduced into Europe relatively recently, after 1850, together with collections of potato species which were imported for breeding purposes. Only a limited part of the genetic variation present in their centre of origin has been introduced into Europe (Folkertsma 1997). From the moment of their introduction onwards, the genetic variation in virulence within and between European nematode populations has been determined predominantly by 1) the genetic structure of the primary founders, 2) random genetic drift and 3) gene flow. Mutation and selection can be excluded as a driving force for the observed variation; the species produce only one generation in a growing season, their multiplication rate is low, the time between generations is 2 to 4 years in normal crop rotation and the active spread of the nematode is limited to several centimeters in the soil. It seems therefore highly unlikely that PCN populations have acquired other virulence characteristics than those already present at the moment of their introduction into Europe. Strategies to obtain broad spectrum resistance against PCN are therefore based on combining a

minimal number of genes with complementary or partially overlapping resistance spectra (Bakker *et al.*, 1993).

Plant disease resistance genes

The majority of plant resistance (R-) genes are located in chromosomal bins containing other disease or insect resistance factors (reviewed in Crute and Pink, 1996). These resistance genes are dominantly inherited, are often involved in resistance processes which are characterized by a hypersensitive response (HR) and are members of multigene families hypothesized to have evolved from common ancestral genes. Most R-loci are characterized by the presence of DNA sequences encoding putative gene products that contain (1) a nucleotide binding site (NBS) and (2) a leucine rich repeat structure (LRR). These structural motifs are known to occur in a large number of resistance gene products; nearly 30 resistance genes from various species have now been cloned and with the exception of two (*Hm1* and *mlo*; Johal and Briggs, 1992; Büschges et al. 1997), these genes are thought to be components of signal transduction pathways (Baker et al. 1997). On the basis of the structural similarity within the motifs of these genes, it is hypothesized that resistance genes are evolutionarily related components of a recognition system (Staskawicz et al. 1995). However, outside these structural motifs, the nucleotide sequences of disease resistance genes are unrelated and several subclasses can be distinguished (Leister et al. 1998). Genes associated with resistance to nematodes in potato are likely to constitute a separate subclass of R-genes. However, the basic architecture hereof has not yet been uncovered. The isolation, characterization and functional analysis of these nematode R-genes remains to be done.

Clustering of R-loci in potato has been reported. One of the large R-loci clusters is on the short arm of potato chromosome *5*. This cluster comprises at least five R-loci: *R1* associated with resistance to *Phytophthora infestans* (Leonards-Schippers et al. 1992), *Nb* associated with HR type resistance to potato virus X (de Jong et al. 1997), *Rx2* associated with an extreme type of resistance to PVX, and *Gpa* and *Grp1* associated with resistance to the PCN (Kreike et al. 1994; Rouppe van der Voort et al. 1998). The recently identified PCN R-locus *Gpa5* is also located within the *Grp1* region (Rouppe van der Voort and Van der Vossen; unpublished data). Additionally, *Gpa6* has been mapped to a region on chromosome 9 on which the homologous region in tomato, *Sw5*, conferring resistance to tomato spotted wilt virus, resides (Rouppe van der Voort and

Van der Vossen; unpublished data).

The *Gpa2* locus

The *Gpa2* locus in potato has been found to be associated with resistance to *G. pallida* populations D383 and D372 (Arntzen et al. 1994). The presence of a single locus in potato which acts specifically to this small cluster of populations indicates that a gene-for-gene relationship underlies this plant-pathogen interaction (Rouppe van der Voort et al. 1997; Bakker et al. 1993). Although, the *Gpa2* locus has previously been mapped on the short arm of chromosome *12* of potato (Rouppe van der Voort et al. 1997a), thusfar no sequence data or precise location were known. The gene was never isolated and no indication as to whether this single sequence would suffice to provide resistance or reduce susceptibility to nematode infection was available.

SUMMARY OF THE INVENTION

The present invention relates to a nucleic acid sequence providing resistance to infection by a phytopathogenic nematode of the *Globodera* species when introduced into a host plant, said host plant prior to introduction being susceptible to infection to the phytopathogenic nematode, said introduction occurring in such a way that said nucleic acid sequence is expressed in the host plant. Furthermore the invention relates to sequences which are homologous to the aforementioned sequence and which, when present in a plant, are able to render said plant resistant to infection by *Globodera* species. More specifically, a sequence according to the invention is preferably that of SEQ ID NO.1 or a homologue thereof. The PCN resistance locus *Gpa2,* when present in a plant such as *Solanum* spp., is capable of conferring to the plant anti-phytopathogenic activity in the form of resistance to *Globodera* species which are known to invade and damage the roots of Solanacae. The invention relates to the *Gpa2* resistance gene of which the DNA sequence is disclosed herein.

The invention also relates to a product encoded by a nucleic acid sequence according to the invention, said product providing nematode resistance activity. Furthermore, the invention relates to genetic constructs, vectors, host cells such as bacterial strains, yeast cells and plant cells comprising a nucleic acid sequence according to the invention. In another aspect, the present invention relates to a genetically transformed plant, preferably of the family Solanacae, especially a genetically

transformed potato plant. Suitably, in a host cell according to the invention, the expression product of the nucleic acid sequence according to the invention, said expression product providing the anti-nematode activity, is produced in an increased amount as compared to the untransformed host cell so as to result in an increased resistance to *Globodera* species. A process for producing a genetically transformed or transfected nematode resistant plant is additionally provided as is a process for isolating or detecting nucleic acid sequences according to the invention, providing nematode resistance of the aforementioned type. A process for diagnosing whether a plant is resistant to *Globodera* species and a process for providing resistance to *Globodera* species to plant material are also disclosed in the present invention. The invention also encompasses a process for producing a polypeptide providing the resistance and a nematocide composition providing said resistance. Antibodies to the polypeptide are also envisaged as embodiments of the invention as is the application thereof in a diagnostic kit for assessing whether a plant is resistant to the aforementioned nematodes. A diagnostic kit according to the invention may also comprise probes and/or primers specific for detection of a nucleic acid sequence providing the resistance.

The present invention relates to oligonucleotides corresponding to a part of a sequence according to the invention or being complementary thereto, with which homologous resistance genes can be identified that confer resistance to *Globodera* species.

DETAILED DESCRIPTION OF THE INVENTION

Definitions

The following definitions are provided for terms used in the description and examples that follow.

- *Nucleic acid*: a double or single stranded DNA or RNA molecule.
- *Oligonucleotide*: a short single-stranded nucleic acid molecule.
- *Primer*: the term primer refers to an oligonucleotide which can prime the synthesis of nucleic acid.
- *Homologous sequence*: a sequence which has at least 70%, preferably 75%, more preferably 80%, most preferably 85% or even 90% sequence identity with the nucleic acid of the invention, whereby the length of the sequences to be compared for nucleic

acids is at least 100 nucleotides, preferably 200 nucleotides and more preferably 300 nucleotides and for polypeptides at least 50 amino acid residues, preferably 75 amino acid residues and more preferably 100 amino acid residues. Homology between the sequences may be as defined and determined by the TBLASTN computer programme for nucleic acids or the TBLASTP computer programme for polypeptides, of Altschul *et al.* (1990), which is in standard use in the art, or, and this may be preferred, the standard program BestFit, which is part of the Wisconsin Package, Version 8, September 1994, (Genetics Computer Group, 575 Science Drive, Madison, Wisconsin, USA, Wisconsin 53711). Alternatively, a homologous sequence refers to a nucleic acid which can hybridize under stringent conditions to the nucleic acid of the invention. Nucleic acid hybridization is a method for detecting related sequences by hybridization of single-stranded nucleic acid probes with denatured complementary target DNA on supports such as nylon membrane or nitrocellulose filters. Nucleic acid molecules that have complementary base sequences will reform the double-stranded structure if mixed in solutions under the proper conditions, even if the target nucleic acid is immobilized on a support. Stringent conditions refer to hybridization conditions which allow a nucleic acid sequence of at least 50 nucleotides and preferably about 200 or more nucleotides to hybridize to a particular sequence at about 65°C in a solution comprising approximately 1 M salt, preferably 6 x SSC or any other solution having a comparable ionic strength, and washing at 65°C in a solution comprising about 0.1 M salt, or less, preferably 0.2 x SSC or any other solution having a comparable ionic strength. These conditions allow the detection of sequences having about 90% or more sequence identity. The person skilled in the art will be able to modify hybridization conditions in order to identify sequences varying in identity between 50% and 90% or more. Binding of the single-stranded nucleic acid probe to a corresponding target nucleic acid may be measured using any of a variety of techniques at the disposal of those skilled in the art.

- *Promoter*: the term "promoter" is intended to mean a short DNA sequence to which RNA polymerase and/or other transcription initiation factors bind prior to transcription of the DNA to which the promoter is functionally connected, allowing transcription to take place. The promoter is usually situated upstream (5') of the coding sequence. In its broader scope, the term "promoter" includes the RNA polymerase binding site as well as regulatory sequence elements located within several hundreds of base pairs, occasionally even further away, from the transcription start site. Such regulatory sequences are, e.g.,

sequences which are involved in the binding of protein factors which control the effectiveness of transcription initiation in response to physiological conditions. The promoter region should be functional in the host cell and preferably corresponds to the natural promoter region of the *Gpa2* resistance gene. However, any heterologous promoter region can be used as long as it is functional in the host cell where expression is desired. The heterologous promoter can be either constitutive or regulatable. A constitutive promoter such as the CaMV 35S promoter or T-DNA promoters, all well known to those skilled in the art, is a promoter which is subjected to substantially no regulation such as induction or repression, but which allows for a steady and substantially unchanged transcription of the DNA sequence to which it is functionally bound in all active cells of the organism provided that other requirements for the transcription to take place is fulfilled. A regulatable promoter is a promoter of which the function is regulated by one or more factors. These factors may either be such which by their presence ensure expression of the relevant DNA sequence or may, alternatively, be such which suppress the expression of the DNA sequence so that their absence causes the DNA sequence to be expressed. Thus, the promoter and optionally its associated regulatory sequence may be activated by the presence or absence of one or more factors to affect transcription of the DNA sequences of the genetic construct of the invention. Suitable promoter sequences and means for obtaining an increased transcription and expression are known to those skilled in the art.

- *Terminator*: the transcription terminator serves to terminate the transcription of the DNA into RNA and is preferably selected from the group consisting of plant transcription terminator sequences, bacterial transcription terminator sequences and plant virus terminator sequences known to those skilled in the art.

- *Nematode*: plant parasitic roundworms of the genus *Globodera,* i.e. *Globodera pallida* and *G. rostochiensis.*

- *Nematode resistance*: to understand the nature of the activity of the *Gpa2* locus in connection with nematode resistance, a brief description of the histopathology of *Solanum* spp. infected with *Globodera* species is hereby given. The infective second-stage larvae hatch and emerge from the cysts and then migrate to and enter roots of susceptible (nonresistant) and resistant potato plants. Before feeding and developing in the root tissue, the nematode induces the formation of multinucleated syncytium. In susceptible potato plants, cessation of feeding by the mature nematode is followed by the

development of cysts breaking out of the root tissue but still clinging to the potato roots. The larvae may survive for a long period in the cysts. In the case of a nematode resistant plant, the number of cysts formed by the adult female nematodes is reduced whereby retardation of the growth of the potato plant is prevented. In accordance herewith, the term "nematode resistance" denotes the characteristic activity in a plant ascribable to a resistance gene, i.e. the capability of the gene products to reduce or prevent the formation of cysts on the roots of plants in particular of Solanacae like e.g. *Solanum* spp.

- *Gene*: the term "gene" is used to indicate a DNA sequence which is involved in producing a polypeptide chain and which includes regions preceding and following the coding region (5'-upstream and 3'-downstream sequences) as well as intervening sequences, the so-called introns, which are placed between individual coding segments (so-called exons) or in the 5'-upstream or 3'-downstream region. The 5'-upstream region comprises a regulatory sequence which controls the expression of the gene, typically a promoter. The 3'-downstream region comprises sequences which are involved in termination of transcription of the gene and optionally sequences responsible for polyadenylation of the transcript and the 3' untranslated region. The term "resistance gene" is a nucleic acid comprising a sequence as depicted in Fig. 3 (SEQ ID NO.3), or part thereof, or any homologous sequence.

- *Resistance gene product*: a polypeptide having an amino acid sequence as depicted in Fig. 3 (SEQ ID NO.1) or part thereof, or any homologous sequence exhibiting the characteristic of providing nematode resistance when incorporated and expressed in a plant.

Scope of the invention

The present invention relates to a nucleic acid sequence providing resistance to infection by a phytopathogenic nematode of the genus *Globodera* when introduced into a host plant, said host plant prior to introduction being susceptible to infection with the phytopathogenic nematode, said introduction occurring in such a way that said nucleic acid sequence is expressed in the host plant. Furthermore the invention relates to resistance sequences which are homologous to the aforementioned sequence and which, when present in a plant, are able to confer to said plant resistance to infection by *Globodera* species. More specifically, a sequence according to the invention is suitably that of SEQ ID NO.1 or a homologue thereof. The PCN resistance locus *Gpa2,* when

present in a plant such as *Solanum* spp., is capable of conferring, to the plant, anti-phytopathogenic activity in the form of resistance to *Globodera* species which are known to invade and damage the roots of Solanacae. The invention relates to the *Gpa2* resistance gene of which the DNA sequence is disclosed herein.

Homologues of the nucleic acid sequence of the abovementioned embodiment of the invention which also provide resistance to *Globodera* species, said homologues being nucleic acid sequences encoding the amino acid sequence of SEQ ID NO.1, are also within the scope of the invention. A homologue of the nucleic acid sequence according to the invention can suitably also provide the resistance when said homologue is a nucleic acid sequence exhibiting more than 70% homology at nucleic acid level with SEQ ID NO. 1. Alternatively the homologue is a nucleic acid sequence exhibiting more than 75% homology at nucleic acid level with SEQ ID NO. 1, preferably exhibiting more than 80% homology at nucleic acid level with SEQ ID NO. 1, more preferably exhibiting more than 85% homology at nucleic acid level with SEQ ID NO. 1. A homologue of the nucleic acid sequence according to the invention, said homologue providing the resistance, can also be a nucleic acid sequence exhibiting more than 90% homology at nucleic acid level with SEQ ID NO.1 and can even be a nucleic acid sequence exhibiting more than 95% homology at nucleic acid level with SEQ ID NO.1. A homologue also providing the resistance can be a nucleic acid sequence capable of hybridising under normal to stringent conditions to the nucleic acid sequence of SEQ ID NO. 1. Naturally another suitable embodiment of a homologue of the sequence according to the invention, also providing the resistance, can be a nucleic acid sequence encoding a deletion, insertion or substitution mutant of the amino acid sequence of SEQ ID NO.1. Such a homologue, also providing the resistance, can be a nucleic acid sequence encoding a deletion, insertion or substitution variant, preferably as occurs in nature, of the amino acid sequence of SEQ ID NO.1. A nucleic acid sequence according to the invention may in addition to any of the embodiments described above or any combinations thereof further comprise at least one intron. Suitable examples of introns and locations thereof are provided in SEQ ID NO.2. A suitable embodiment of the nucleic acid sequence according to the invention is the genomic insert of pBINRGH2 as disclosed in the examples. A nucleic acid sequence according to the invention is suitably identical to that present in the genetic material of a species of the Solanacae family, preferably a species of the genus *Solanum*. More specifically, such sequences can be

found on and are preferably identical to those present in the genome of potato on chromosomes 4, 5, 7, 9, 11 or 12. More specifically, the nucleic acid sequence is identical to that present in the genome of potato at the *Gpa2* locus. Obviously, a fragment of any of the above mentioned embodiments exhibiting the characteristic of providing the resistance falls within the scope of the invention.

According to the present invention, a DNA region comprising the PCN R-locus *Gpa2* has been isolated from a potato plant harbouring a wild *Solanum* genomic introgression segment possessing resistance against nematode infection. This resistance, which appears to be very effective in PCN control, is not present in most cultivated potato cultivars. Therefore, one object of the present invention is to provide plants, specifically *Solanum* spp., which have the features of cultivated plants, with anti-phytopathogenic activity in the form of resistance to *Globodera* species. Thus the present invention relates to a DNA segment comprising the *Gpa2* locus of about 200 kb comprising one or several genes, the gene product or gene products thereof being capable of conferring to the plant resistance to nematodes of the *Globodera* species.

Another aspect of the present invention is a nucleic acid comprising the *Gpa2* resistance gene, the nucleic acid having the sequence of all or part of the sequence depicted in Fig. 3 (SEQ ID NO.3) or any homologous sequence, including (where appropriate) both coding and/or noncoding regions and providing nematode resistance upon expression thereof in a plant. In a preferred embodiment the *Gpa2* gene comprises the deduced coding sequence provided in Fig. 3 (SEQ ID NO.1) or any homologous sequence, preceded by a promoter region and followed by a terminator sequence.

As described in the invention, the nucleic acid sequence according to the invention possesses very valuable features with respect to anti-nematode activity. Thus, the DNA region comprising the nucleic acid sequence according to the invention encoding a polypeptide conferring/evoking the anti-nematode activity as defined above, can be used for the construction of genetically modified hosts having an increased resistance to nematodes as compared to untransformed hosts. The nucleic acid region according to the invention is thus capable of being inserted into the genome of a host plant, which in itself is susceptible to infection by a nematode, in such a way that the nucleic acid sequence is expressed, thereby conferring to the host plant resistance to infection by a phytopathogenic nematode. Thus, another aspect of the present invention relates to a genetic construct consisting of the nucleic acid sequence according to the

invention which genetic construct can then be used to genetically transform a host, e.g. a plant such as a cultivated plant, in such a way that it becomes resistant to nematodes.

A genetic construct comprising a nucleic acid sequence according to any of the embodiments described above, said sequence being operably linked to a regulatory region for expression, falls within the scope of the invention. Accordingly, the present invention relates to a genetic construct comprising

1) a promoter functionally connected to

2) a nucleic acid region as defined according to the present invention

3) a transcription terminator functionally connected to the nucleic acid sequence.

Preferably, the regulatory region of a genetic construct according to the invention is a *Gpa2* regulatory region. Such a regulatory region can by way of example correspond to that present in the sequence of nucleotides 1-4874 of SEQ ID NO.3. The regulatory region can suitably even correspond to that of nucleotides 1-4874 of SEQ ID NO.3. The regulatory region preferably comprises a promoter functionally connected to the nucleic acid sequence as defined in any of the embodiments above or in the examples, said promoter being able to control the transcription of said nucleic acid sequence in a host cell, preferably in a plant cell.

The genetic construct may be used in the construction of a genetically modified host in order to produce a host showing an increased anti-nematode activity and thus an increased resistance towards nematodes. It will be understood that a large number of different genetic constructs as defined above may be designed and prepared. Without being an exhaustive list, elements of the genetic constructs which may be varied are the number of copies of each of the nucleic acid sequences of the genetic construct, the specific nucleotide sequence of each of the nucleic acid sequences, the type of promoter and terminator connected to each nucleic acid sequence, and the type of any other associated sequences. Thus, genetic constructs of the present invention may vary within wide limits.

The invention also relates to DNA constructs comprising the regulatory sequences, and more preferably the promoter region of the *Gpa2* resistance gene in conjunction with a structural gene sequence heterologous to said regulatory sequences.

A vector which carries a nucleic acid according to any of the embodiments disclosed above or in the examples or a genetic construct according to any of the embodiments disclosed above or in the examples also falls within the scope of the

invention. Preferably such a vector is capable of replicating in a host organism. The vector may either be one which is capable of autonomous replication, such as a plasmid, or one which is replicated with the host chromosome such as a bacteriophage or integrated into a plant genome. For production purposes, the vector is an expression vector capable of expressing the nucleic acid sequence according to the invention in the organism chosen for the production. Suitable cloning vectors, transformation vectors, expression vectors, etc..., are well known to those skilled in the art. A vector according to the invention is constructed to function in a host organism selected from the group consisting of a micro-organism, plant cell, plant, seed, seedling, plant part and protoplast. A host cell capable of resulting in a plant is preferred and suitably the host organism is selected from the group consisting of a plant, plant cell, plant part, seed, seedling and protoplast.

In a still further aspect, the present invention relates to a host organism which carries and which is capable of replicating or expressing an inserted nucleic acid region of the invention. Such a host organism is preferably selected from the group consisting of a micro-organism, plant cell, plant, seed, seedling, plant part and protoplast, harbouring a vector and/or a genetic construct as defined above. The term "inserted" indicates that the nucleic acid region has been inserted into the organism or an ancestor thereof by means of genetic manipulation, in other words, the organism may be one which did not naturally or inherently contain such a nucleic acid region in its genome, or it may be one which naturally or inherently contains such a nucleic acid region, but in a lower number so that the organism with the inserted nucleic acid region has a higher number of such regions than its naturally occurring counterparts. The nucleic acid region carried by the organism may be part of the genome of the organism, or may be carried on a genetic construct or vector as defined above which is harboured in the organism. The nucleic acid region may be present in the genome or expression vector as defined above in frame with one or more second nucleic acid regions encoding a second gene product or part thereof so as to encode a fusion gene product. The organism may be a higher organism such as a plant, or a lower organism such as a micro-organism. A lower organism such as a bacterium, e.g. a gram-negative bacterium such as a bacterium of the genus *Escherichia,* e.g. *E. coli,* or a yeast such as of the genus *Saccharomyces,* is useful for producing a recombinant polypeptide as defined above. The recombinant production may be performed by use of conventional techniques, e.g. as described by Sambrook et

al. (1990). Also, the organism may be a cell line, e.g. a plant cell line. Most preferably, the organism is a plant, i.e. a genetically modified plant such as will be discussed in further detail below. As mentioned above, the genetic construct is preferably to be used in modifying a plant. Accordingly, the present invention also relates to a genetically transformed plant comprising in its genome a genetic construct as defined above. The genetically transformed plant has an increased anti-nematode activity compared to a plant which does not harbour a genetic construct of the invention, e.g. an untransformed or natural plant or a plant which has been genetically transformed, but not with a genetic construct of the invention. Normally a constitutive expression of the gene products encoded by the genetic construct is desirable, but in certain cases it may be preferable to have the expression of the gene products encoded by the genetic construct regulated by various factors, for example by factors such as temperature, pathogens, and biological factors. The genetically transformed plant is obtained by introducing the nucleic acid sequence according to the invention within the genome of said plant having a susceptible genotype to nematodes, using standard transformation techniques. It will be apparent from the above disclosure, that the genetically transformed plant according to the invention has an increased resistance to nematodes as compared to plants which have not been genetically transformed according to the invention or as compared to plants which do not harbour the genetic construct as defined above. In a further aspect, the present invention relates to seeds, seedlings or plant parts obtained by growing the genetically transformed plant as described above or by genetically transforming a plant cell and generating said part. It will be understood that any plant part or cell derivable from a genetically transformed host of the invention is to be considered within the scope of the present invention.

A process for producing a genetically transformed host organism having increased resistance to *Globodera* species as compared to the host organism prior to the transformation, said process comprising transferring a genetic construct and/or a vector according to any of the embodiments disclosed above and in the examples into the host organism so that it's genetic material comprises the genetic construct and/or vector and subsequently regenerating the host organism into a genetically transformed plant part is also a part of the invention. The host organism may be selected from the group consisting of a plant cell, plant, seed, seedling, plant part and protoplast of the plant type to be rendered resistant and may subsequently be regenerated to a plant. Preferably, the

nematodes against which resistance is provided are selected from the group consisting of *Globodera* species, more specifically *Globodera rostochiensis* and *Globodera pallida*. The host organism which is to be transformed is selected from a plant type of the family Solanacae, preferably a *Solanum* spp. Plants of the species *Solanum tuberosum*, comprising commercial potato cultivars, are preferred as this is a particular problem area for the commercial growers of such plants.

In accordance with well-known plant breeding techniques it will be understood that the production of a genetically transformed plant may be performed by a double transformation event (introducing the genetic construct in two transformation cycles) or may be associated with use of conventional breeding techniques. Thus, two genetically modified plants according to the present invention may be the subject of cross breading in order to obtain a plant which contains the genetic construct of each of its parent plants.

Additionaly, the present invention also relates to the resistance gene product which is encoded by the nucleic acid sequence according to the invention and which has the deduced amino acid sequence provided in Fig. 3 (SEQ ID NO.1). Thus a polypeptide having an amino acid sequence provided in SEQ ID NO.1 and also a homologue of said amino acid sequence, said homologue being a substitution, insertion or deletion mutant conferring nematode resistance against *Globodera* species, form embodiments of the invention. A polypeptide according to the invention is encoded by a sequence according to any of the embodiments described above or in the examples. A process for producing such polypeptides having an amino acid sequence provided in SEQ ID NO.1, or a homologue of said amino acid sequence, said homologue being a substitution, insertion or deletion mutant possessing resistance to *Globodera* species, said process comprising the expression of the nucleic acid sequence or genetic construct according to any of the embodiments according to the invention and optionally isolating said polypeptide, said expression occurring in a host organism according to the invention, is also covered by the invention. A process comprising an isolation step of the polypeptide in a manner known *per se* for polypeptide isolation from cell culture or from the host organism itself is also covered.

A nematicide composition comprising as active component a polypeptide according to the above or produced according to the process described or a host organism expressing such a polypeptide in a formulation suitable for application as

nematicide to a plant and optionally comprising other ingredients required for rendering the polypeptide suitable for application as a nematicide, also falls within the scope of the invention. Preferably such a nematicide composition comprises the polypeptide in a slow release dosage form. It is quite efficient if such a nematicide composition is formulated and packaged comprising instructions for application as nematicide.

Antibodies may be raised against any purified resistance gene product according to the invention by any method known to those skilled in the art (for an overview see "Immunology - 5th Edition" by Roitt, Male: Pub 1998-Mosby Press, London). Such antibodies can be used for the detection of the gene product.

Another aspect of the invention relates to nucleic acid sequences comprising at least 16 contiguous nucleotides corresponding to or complementary to the *Gpa2* sequence, with the proviso that when such a nucleic acid comprises part or all of the NBS encoding sequence, the nucleic acid also comprises at least one codon of a 5' and/or 3' overhanging portion corresponding to the respective 5' and/or 3' adjacent amino acids of parts of the NBS sequence of the *Gpa2,* with the following sequence, GGIGKTT or GGLPLA (see Table 4). Preferably, the *Gpa2* sequence is comprised within the sequence of SEQ ID NO.1, 2 or 3. The sequence length is preferably at least 50 nucleotides, preferably more than 100 nucleotides rendering it suitable for use as a probe in a nucleic acid hybridization assay. Oligonucleotides complementary to one strand of the *Gpa2* sequence or part thereof, can be used as labeled hybridization probes in a Southern hybridization procedure or as primers in an amplification reaction such as the polymerase chain reaction (PCR), for the screening of genomic DNA or cDNA, or constructed libraries thereof, for the identification and isolation of homologous genes. Homologous genes that are identified in this way and which encode a gene product that is involved in conferring reduced susceptibility or resistance to a plant against pathogens, such as nematodes of the genus *Globodera*, are comprised within the scope of the invention. Suitable embodiments can be selected from any of the sequences SEQ. ID. No.4, 5, 6 and/or 7.

A diagnostic kit for assessing the presence of nematode resistance in a plant to infection by a phytopathogenic nematode of the genus *Globodera*, said kit comprising at least one nucleic acid defined above as a probe or primer, for screening of nucleic acid from a plant or plant part to be tested and/or comprising an antibody as defined above, is also comprised within the scope of the invention.

The invention also covers a process for isolating or detecting a nucleic acid sequence according to the invention providing nematode resistance as described above and in the examples, said process comprising the screening of genomic nucleic acid of a plant with said nucleic acids or a fragment thereof as probe or primer, said probe or primer being at least 16 nucleotides in length, the identification of positive clones which hybridize to said probe or primer and the isolation of said positive clones and the isolation of the nucleic acid sequence therefrom. Such a process comprises screening genomic nucleic acid of a plant, preferably such a process comprises the screening of a genomic library of a plant with a nucleic acid sequence according to SEQ ID NO 3 or a fragment thereof as probe or primer, said probe being at least 16 nucleotides in length. Alternatively such a process comprises the screening of a cDNA library of a plant with the coding portion of a nucleic acid sequence according to the invention providing the nematode resistance, or a fragment thereof as probe or primer, said probe or primer being at least 16 nucleotides in length. Preferably, for the screening of a cDNA library of a plant, the coding portion of a nucleic acid according to SEQ ID NO.1 or a fragment thereof is used as probe or primer. The probe or primer can be comprised within the sequence of SEQ ID NO 1, SEQ ID NO 2 or SEQ ID NO 3. The above processes can use a nucleic acid amplification reaction such as PCR in conjunction with at least one primer corresponding to or being complementary to the nucleic acid sequence according to the invention providing the nematode resistance, or a fragment thereof, said primer being at least 16 nucleotides in length. The primer can be complementary to the nucleic acid sequence of SEQ ID NO.1, SEQ ID NO.2 or SEQ ID NO.3 or a fragment thereof, said primer being at least 16 nucleotides in length. A probe or primer in such a process comprises a nucleic acid sequence encoding the amino acid sequence of a part or all of the NBS sequence of *Gpa2*. Suitable examples of primers comprising a nucleic acid sequence encoding the amino acid sequence of a specific part or all of the NBS sequence of *Gpa2* are given below (see Table 4). For reasons of specificity, the process can comprise application of a primer comprising at least part of the NBS sequence of *Gpa2* and at least one codon of a 5' and/or 3' overhanging portion corresponding to the respective 5' and/or 3' adjacent amino acids of the previously specified NBS sequence of *Gpa2*. An example of such a primer comprises the specified part of the NBS sequence of *Gpa2* and at least one codon of a 5' and/or 3' overhanging portion corresponding to the respective 5' and/or 3' adjacent amino acids of the NBS sequence of *Gpa2* of SEQ ID

NO.1. Preferably, said primers correspond to a sequence selected from SEQ ID NO.4, SEQ ID NO.5, SEQ ID NO.6 and/or SEQ ID NO.7.

A process for diagnosing whether a plant is resistant to a phytopathogenic *Globodera* species, said process comprising the detection of the presence of a nucleic acid sequence providing nematode resistance as defined in any of the embodiments according to the invention, the presence of a genetic construct according to any of the embodiments according to the invention, the presence of a vector according to any of the embodiments according to the invention or the presence of a polypeptide as defined according to the invention, in the genetic material of plant material of a plant to be tested falls within the scope of the invention. Combinations of detection of the various elements are also covered. The nucleic acid sequence and the polypeptide being detected can be naturally present in the plant to be tested or can have been introduced via genetic engineering. A process for diagnosis according to the invention can comprise any of the nucleic acid sequence detection processes already described above as embodiments of the invention. More specifically the process can comprise applications of the diagnostic kit described according to the invention in an analogous manner to application of other nucleic acid assay kits comprising probes or primers or antibody known in the art. Suitably such a kit according to the invention will be provided with the appropriate instructions for application thereof. Amplification reactions of nucleic acid , use of probes in Southern analysis and use of antibodies in immunoassays are suitable examples of applications known in the art.

Another process within the scope of the invention is a process for providing resistance to a phytopathogenic *Globodera* species to plant material, said process comprising the introduction into the plant genome of a nucleic acid sequence providing nematode resistance as defined in any of the embodiments according to the invention, a genetic construct according to any of the embodiments according to the invention, a vector according to any of the embodiments according to the invention in the genetic material of plant material of a plant to be tested and thereby producing a transformed plant cell, plant propagating material, plant part or plant. Such introduction of genetic material should result in a transformed host with the introduced genetic material stably present in the host such that replication of said host is possible with said genetic material. Such a process may further comprise regenerating the resulting transformed or transfected plant cell, plant propagating material or plant part. The process of

introduction of the genetic material can occur as commonly described in the art for introduction of genetic material into the appropriate host type.

The nucleic acid sequence comprising the resistance as provided by the present invention has numerous applications of which some are described herein but which are not limiting to the scope of the invention.

The present invention is further described in detail below, whereby the map-based cloning strategy used to isolate the *Gpa2* resistance gene of the invention is explained. The strategy to isolate the *Gpa2* gene was as follows:

1) genetic fine mapping of the *Gpa2* locus;
2) construction of a BAC contig spanning the *Gpa2* locus;
3) identification of candidate resistance gene homologues (RGH);
4) complementation analysis.

The *Gpa2* locus was initially mapped on chromosome 12 using information on the genomic positions of 733 known AFLP markers (Rouppe van der Voort *et al.*, 1997a and 1997b). By use of RFLP probes, *Gpa2* was mapped more precisely between markers GP34 and CT79 on the distal end of chromosome 12 (Rouppe van der Voort *et al.*, 1997a), a 6 cM genetic interval that was previously shown to harbour the potato virus X (PVX) resistance gene *Rx*1 (Fig. 1; Bendahmane *et al.*, 1997). Cosegregation of *Gpa2* and *Rx*1 in the tetraploid *Rx*1 mapping population (S1-Cara) and a diploid *Gpa2* mapping population (F1SHxRH) confirmed the assumed linkage between the two genes. The S1-Cara recombinants initially chosen to confirm this linkage delimited the *Gpa2* interval between markers IPM3 and IPM5 (Fig. 2; Bendahmane *et al.* 1997).

Fine mapping of the *Gpa2* locus was subsequently carried out using cleaved amplified polymorphic sequence (CAPS; Konieczny and Ausubel, 1993) markers derived from the IPM3-IPM5 interval, all of which were initially developed for the cloning of *Rx*1 (Fig. 1). 2,788 S1-Cara genotypes were assayed for recombination events in the IPM3-IPM5 region. In addition, 598 F1SHxRH genotypes were subjected to a GP34/IPM5 marker screening as marker IPM3 was not informative in population F1SHxRH. Plants with recombination events between these markers were subsequently tested for all markers available in the IPM3-IPM5 region as well as for *Gpa2* resistance. This analysis showed that *Gpa2* is located between markers IPM4c and 111R (Fig. 2). Among the 2,788 S1-Cara genotypes and 598 F1SHxRH genotypes tested, only one genotype, S1-761, was identified in which a recombination event had occurred between

Gpa2 and marker IPM4c (Fig. 2B). On the other side of *Gpa2*, genotype S1-B811 could be used to identify marker 111R as a flanking marker for the *Gpa2* interval (Fig. 2B).

Four BAC clones, BAC77, BAC45, BAC221 and BAC111, which map to the 0.06 cM IPM4c-111R genetic interval harbouring the *Gpa2* locus, were isolated from a BAC library prepared from a progeny of a selfed cv. Cara (Fig. 1C). However these four BAC clones did not completely cover the *Gpa2* interval. Screening of the Cara BAC library with CAPS markers 77R and 45L (Fig. 1B) did not lead to the identification of Cara BAC clones that spanned the region between markers 77R and 45L. A second BAC library was constructed from the diploid potato genotype SH83-92-488 (SH83). Screening of the SH83 potato BAC library with CAPS markers 77R and 45L did result in the identification of such a BAC clone (SHBAC43). In this way a contiguous physical map of the IPM4c-111R *Gpa2* interval was constructed comprising SHBAC43, BAC45, BAC221a and BAC111 (see Fig. 2C). Restriction analysis of this BAC contig delimited the physical size of the *Gpa2* locus of approximately 200 kb.

As the size of the *Gpa2* locus was still too large for direct localization of the *Gpa2* resistance gene by complementation analysis, BAC clones SHBAC43, BAC45, BAC221a and BAC111 were analysed for the presence of R-gene homologous sequences. Despite the general lack in DNA sequence conservation between R-genes, there are a few conserved protein motifs in the NBS region present in many of these genes. Leister *et al* (1996) has shown that it is possible to amplify resistance gene like sequences from potato using degenerate primers based on these homologous regions. Using degenerate primers RG1 and RG2 (Aarts *et al.*, 1998), whose sequences are based on the conserved P-loop and domain 5 region of the NBS in the N, L6 and RPS2 R-genes (Whitham *et al.*, 1994; Lawrence *et al.*, 1995; Bent *at al*, 1994 and Mindrinos *et al.*, 1994) a DNA fragment of the expected size (approximately 530 bp) was amplified from BAC221a. Southern analysis of *Eco*RI restricted DNA of SHBAC43, BAC45, BAC221a and BAC111 using the amplified PCR fragment from BAC221a as a probe, identified two copies of this R-gene like sequence on SHBAC43, one single copy on BAC221a and one copy on BAC111 (Fig. 2D). Subsequent sequence analysis of the complete inserts of these BAC clones showed that the previously identified R-gene like sequences on the BAC clones belonged to putative resistance gene homologues (RGHs). Three of these RGH sequences were designated to be candidates for the *Gpa2* gene and selected for complementation analysis; RGH1 on BAC221a, RGH2 on BAC111 and

RGH3 on SHBAC43. A fourth RGH identified on SHBAC43 contained marker IPM4c and therefore lies outside of the *Gpa2* interval (see Fig. 2C and 2D).

Genomic fragments of approximately 11 kb, 10.3 kb and 5.5 harbouring RGH1, RGH2 and RGH3, respectively, were subcloned from the BAC inserts into the plant transformation vector pBINPLUS (Van Engelen *et al.*, 1995) and transferred to a susceptible potato genotype using standard transformation methods. Roots of *in vitro* grown primary transformants were tested for PCN resistance as described in Example 1. This *in vitro* resistance assay revealed that the 10.3 kb genomic insert harbouring RGH2 was able to complement the susceptible phenotype. RGH2 was therefore designated the *Gpa2* gene, the DNA sequence which is provided in Fig. 3.

The following examples provide a further illustration of the present invention which is nevertheless not limited to these examples.

EXAMPLES

EXAMPLE 1: ASSESSING NEMATODE RESISTANCE

A. *In vivo* resistance assay

Eggs and second stage juveniles (J_2) are obtained by crushing cysts which have been soaked in tap water for one week. The egg/J_2 suspension is poured through a 100 µm sieve to remove debris and cystwalls. Before inoculation, three to four week old potato stem cuttings are transferred from a peat mixture to 900 gram pots containing a mixture of silversand and a sandy loam fertilized with Osmocote (N-P-K granulates). Subsequently, plants are inoculated with nematodes to a final density of 5 egg/J_2 per gram soil. Of each plant genotype, three replicates per nematode source are inoculated. Six replicates of the parental clones as well as resistant and susceptible standards are included for resistance tests with each nematode source. Resistant standards are *Solanum tuberosum* cv. Multa (resistant to *G. pallida* D383), *S. vernei* hybrid 58.1642/4 (resistant to *G. rostochiensis* line Ro_1-19) and *S. vernei* hybrid 62-33-3 (resistant to both D383 and Ro_1-19). The susceptible standard is *S. tuberosum* cv.Eigenheimer. Plants are arranged in a randomized block design and grown in a greenhouse with 15°C and 25°C as minimum and maximum temperature, respectively.

After three months, cysts are recovered from the soil with a Fenwick can

(Fenwick 1940) and the size of the root systems is judged on a scale of 0 to 3. Resistance data of a genotype are only recorded when at least two well-rooted plants of this genotype are available. The mean cyst numbers developed per genotype are standardized using a $\log_{10}(x + 1)$ transformation and then subjected to SAS Ward's minimum variance cluster analysis (SAS Institute Inc., Cary NC, USA). On the basis of this analysis the plants are devided into a resistant, an unassigned or a susceptible class.

B. *In vitro* resistance assay

Alternatively, the resistance assay is carried out on sterile tissue culture plants in agar. Two or three nodia from each *in vitro* grown (transgenic) potato plant are grown on solidified B5 medium (Gamborg *et al.* 1968) with 0.5% Phytagel™ (Sigma) and 2% sucrose for one week (25°C and 16 hr light regime). Each new root tip (on average 2 per nodium) is then inoculated with 15 sterilized second stage juveniles. Preparation of inoculum is esentially as described by Heungens *et al.* (1995) with slight modifications. Cysts are collected in a modified 20 ml syringe with a 22 µm nylon mesh and surface sterilized in 90% ethanol for 15 sec followed by an 8 min wash in 1.3% (w/v) commercial bleach. To remove excess bleach, the cysts are washed three times in sterile tap water for 5 min and incubated in sterile tap water for 3 days at 20°C in the dark. Cysts are then transferred to filter sterile potato root differentiate (PRD) and left to hatch for 5 days at 20°C in the dark. Second stage juveniles are subsequently transferred to a 5 µm sieve-syringe and incubated first in 0.5% (w/v) streptomycine-penicilline G solution for 20 min, then in 0.1% (w/v) ampicillin-gentamycin solution for 20 min and finally in 0.1% chlorhexidin solution for 3 min. After three 5 min wash steps in sterile tap water the second stage juveniles are suspended in the required volume (sterile tap water) for inoculation. The petridishes with the inoculated root tips are incubated in the dark at 20°C. After 5-6 weeks the level of infection is determined by counting the number of female nematodes formed on the roots.

EXAMPLE 2: COSEGREGATION OF *Gpa2* NEMATODE RESISTANCE AND *Rx1* VIRUS RESISTANCE.

The *Gpa2* locus was initially mapped to chromosome 12 using information on the

genomic positions of 733 known AFLP markers (Rouppe van der Voort *et al.*, 1997a and 1997b). By use of RFLP probes, *Gpa2* was mapped more precisely between markers GP34 and CT79 on the distal end of chromosome 12 (Fig. 2A; Rouppe van der Voort *et al.*, 1997a), a 6 cM genetic interval that was previously shown to harbour the potato virus X (PVX) resistance gene *Rx*1 (Bendahmane *et al.*, 1997).

To confirm the assumed linkage between *Gpa2* and *Rx*1 (Rouppe van der Voort et al. 1997), a pilot experiment was carried out in which the segregation of both genes was followed in two different mapping populations; a tetraploid (2n = 4x = 48) mapping population derived from a selfing of potato cv. Cara (S1-Cara), initially constructed for fine mapping of *Rx*1 (Bendahmane et al. 1997), and the diploid (2n = 2x = 24) *Gpa2* mapping population derived from a cross between the diploid potato clones SH83-92-488 and RH89-039-16 (F1SHxRH; Rouppe van der Voort *et al.*, 1997a and 1997b). Potato genotypes Cara and SH have the wild accession *Solanum tuberosum* spp. *andigena* CPC 1673 in common.

The tests for *Gpa2* and *Rx*1 resistance were performed on the parental genotypes Cara, SH83 and RH89, four S1 genotypes which were recombined in a 1.21 cM interval between markers GP34 and IPM5 (Fig. 1B; Bendahmane *et al.* 1997) and two F1SHxRH genotypes which harboured cross-over events in a 6 cM interval between markers GP34 and CT79 (Rouppe van der Voort *et al.* 1997). The PVX resistance assay was carried out using a cDNA of the PVX_{CP4} isolate (Goulden *et al.* 1993). Potato plants were graft-inoculated with scions of *Lycopersicon esculentum* cvs. Ailsa Craig or Money Maker systemically infected with PVX_{CP4}. Northern blots were prepared from total RNA isolated from newly formed potato shoots 3-4 weeks post-inoculation (Bendahmane *et al.* 1997). Extreme PVX resistance or susceptibility was determined by the presence or absence of a hybridization signal on Northern blots probed with ^{32}P-labelled cDNA of PVX_{CP4} (Chapman *et al.* 1992). Three replicates per genotype were assayed. For the *Gpa2* assay *G. pallida* population D383 was used (Rouppe van der Voort *et al.* 1997a). The nematode resistance assay was performed as described in Example 1A. Nematode population Rookmaker with different virulence characteristics as population D383 (Bakker *et al.* 1992) was used to confirm the specificity of *Gpa2* resistance in tested plants.

The resistance tests showed a clear reduction in the number of cysts of *G. pallida* population D383 on genotypes which were resistant to PVX_{CP4}. The number of cysts

developed on the resistant S1-Cara genotypes appeared to be slightly higher than the number of cysts found on the resistant genotypes of population F1SHxRH. However, a considerable reduction in size of these cysts was observed as compared to the cysts developed on a susceptible genotype. This observation was corroborated after comparing the number of eggs per cyst developed on resistant and susceptible genotypes. Average cyst contents were determined from at least 30 cysts (if possible) and subjected to a *t*-test. A significant difference (at $P < 0.05$) was found between the average number of eggs per cyst developed on Cara, SH83 and cv. Multa (resistant control), and average egg contents per cysts recovered from genotype S1-350, RH89 and cv. Eigenheimer (susceptible control). Resistance tests using *G. pallida* population Rookmaker show that cv. Cara is susceptible to this nematode population, indicating a specificity for the *G. pallida* resistance in population S1-Cara.

Although limited numbers of S1-Cara and F1SHxRH genotypes were tested for resistance to *G. pallida* population D383 and PVX respectively, based on the position of the crossover events in the tested plants it could be concluded that *Gpa2* and *Rx*1 cosegregate in both mapping populations (with a maximum probability of $P = 1/64$ that the observed linkage could be explained by chance). The tested S1-Cara recombinants were previously used to delimit the *Rx*1 interval between markers IPM3 and IPM5 (Bendahmane *et al.* 1997). Cosegregation of *Gpa2* with *Rx*1 indicates therefore that *Gpa2* also resides in this region (Fig. 2A).

EXAMPLE 3: ISOLATION OF CARA BAC CLONES AND PRODUCTION OF CAPS MARKERS DERIVED FROM THE *Rx*1/*Gpa*2 LOCUS (according to the unpublished article in preparation of Kanyuka, K., Bendahmane, A., Rouppe van der Voort, J.N.A.M., van der Vossen, E.A.G. and Baulcombe, D.C. Mapping of intra-locus duplications and introgressed DNA: aids to map-based cloning of genes from complex genomes illustrated by analysis of the Rx locus in tetraploid potato).

Construction of a Cara BAC library

In order to clone the *Rx*1 locus, a BAC library of 160,000 clones was prepared from plant SC-781 which is a progeny of selfed cv Cara carrying *Rx*1 in the duplex condition (*Rx,Rx,rx,rx*). High molecular weight DNA was prepared in agarose plugs from potato protoplasts essentially as described in Bendahmane *et al.* (1997). The agarose plugs

were dialysed three times for 30 min against TE buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA), once at room temperature and twice at 4°C. The plugs were then equilibrated in *Hin*dIII buffer (10 mM Tris-HCl, 10 mM $MgCl_2$, 50 mM NaCl, 1 mM DTT, pH 7.9) twice on ice for 1 h. Half of each plug (~5 µg of DNA) was transferred to a test tube containing 360 µl of *Hin*dIII buffer and 10-15 units of *Hin*dIII restriction enzyme. The enzyme was allowed to diffuse into a plug at 4°C for 1 h and the digestion was carried out at 37°C for 30 min. The reaction was stopped by adding 1 ml of 0.5 M EDTA and plugs were immediately loaded into a 1% low melting point agarose gel and subjected to contour-clamped homogeneous electric fields (CHEF; Chu, 1989) electrophoresis in a CHEF DR II system (Bio-Rad Laboratories, USA) in 0.5 X TBE buffer (45 mM Tris-borate pH 8.0, 1 mM EDTA) at 150 volts for 10 h at 4°C and constant pulse time of 5 sec or 8 sec. Compression zones containing the DNA fragments of 100 kb or 150 kb were excised from the gel and dialysed against 30 ml TE in a 15 cm Petri dish for 2 h at 4°C. Dialysed agarose slices were then transferred to a 1.5 ml test tube, melted at 70°C for 10 min and digested with 1 unit of GELASE (Epicentre Technologies, USA) per 100 mg of agarose gel for 1 h at 45°C.

The size selected potato DNA (25-50 ng) was ligated to 25-50 ng of *Hin*dIII-digested and dephosphorylated pBeloBAC11 (Shizuya *et al.,* 1992) using 400 to 800 units of T4 DNA LIGASE (New England BioLabs, USA) at 16°C for 24 hours in a total volume of 50 µl. The ligation products were dialysed against 1 X TE using 0.025 µm MF-MILLIPORE MEMBRANE FILTER (Millipore, UK) at 4°C for 2 h and 30 min at room temperature using the "drop dialysis" method of Maruzyk and Sergeant (1980).

Transformation of *E. coli* DH10B cells was carried out by electroporation using a BRL CEMI-PORATOR SYSTEM (Life Technologies Ltd, UK). To 20 µl of electro-competent cells, 0.5-3 µl of ligation mixture was added. After electroporation, *E. coli* cells were mixed with 1 ml SOC medium (0.5% yeast extract, 2% tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose) and incubated at 37°C for 1 h with gentle shaking (80 rpm). The cells were spread on Luria broth (LB) agar plates containing chloramphenicol (12.5 lg/ml), 5-bromo-4-chloro-3-indolyl-ß-D-galactoside (Xgal) (40 lg/ml) and isopropyl-1-thio-ß-D-galactoside (IPTG) (0.12 mg/ml). Plates were incubated at 37°C for 24 hours. DNA from the compression zones of 100 and 150 kb led to clones with an average insert size of 100 kb and a transformation efficiency of approximately 1000 and 150 white colonies per ll ligation mixture,

respectively. Approximately 92000 white colonies from these ligations were picked individually into 384 well microtiter plates (Genetix, UK) containing LB freezing buffer (36 mM K_2HPO_4, 13.2 mM KH_2PO_4, 1.7 mM citrate, 0.4 mM $MgSO_4$, 6.8 mM $(NH_4)_2SO_4$, 4.4 % V/V glycerol, 12.5 µg/ml chloramphenicol in LB medium), grown overnight at 37°C and stored at -80°C. Another 100 bacterial pools containing ~500-1000 white colonies each (these pools also contained approximately 500-1500 blue bacterial colonies with an empty pBeloBAC11) were prepared by scraping the colonies from agar plates into the LB medium containing 18% glycerol and 12.5 µg/ml chloramphenicol using a sterile glass spreader. These pools were also stored at -80°C.

Screening of the Cara BAC library with markers IPM3, IPM4 and IPM5 and isolation of BAC clones derived from the Rx1/Gpa2 locus

The Cara BAC library was initially screened with CAPS markers IPM3, IPM4 and IPM5 corresponding to the AFLP markers PM3, PM4 and PM5 flanking the *Rx*1 locus (Bendahmane *et al.*, 1997). This was carried out as follows. For the first part of the library of 92,160 clones stored in 384 well microtiter plates the plasmid DNA was isolated using the standard alkaline lysis protocol (Heilig *et al.*, 1997) from pooled bacteria of each plate to produce 240 plate pools. Aliquots of these plate pools were combined to prepare 26 'superpools' containing DNA from 9 plate pools, and one superpool containing DNA from 6 plate pools. To identify individual BAC clones carrying the CAPS markers the superpools and then the corresponding plate pools were screened. Once an individual plate had been identified the clones corresponding to each of the 24 columns of the positive plate were grown for 3-4 h at 37°C in LB medium and PCR was carried out on 3 µl of bacteria. After identification of a positive column a colony PCR on each of the corresponding 16 colonies of this column was carried out leading to identification of a single positive BAC clone.

For the second part of the library, which is stored as one hundred pools of approximately 1000 clones, plasmid DNA was isolated from each pool of clones using the standard alkaline lysis protocol and PCR was carried out to identify positive pools. Bacteria corresponding to positive pools were diluted, plated on LB agar plates and subsequently colony hybridisation was carried out as described in Sambrook *et al.* (1989) using ^{32}P-labelled DNA probes corresponding to the CAPS markers. PCR with the corresponding CAPS primers was used to distinguish between hybridising colonies

carrying the markers previously mapped to homologues located elsewhere in the genome and those derived from the *Rx*1 locus.

Positive BAC clones were analysed by isolating plasmid DNA from 5 ml overnight cultures (LB medium supplemented with 12.5 mg/ml chloramphenicol) using the standard alkaline lysis miniprep protocol (Engebrecht *et al.*, 1997) and resuspended in 50 µl TE. Plasmid DNA (10 µl) was digested with *Not*I for 3 h at 37°C to free the genomic DNA from the pBeloBAC11 vector. The digested DNA was separated by CHEF electrophoresis in a 1% agarose gel in 0.5 X TBE at 4°C using a BIO-RAD CHEF DR II system (Bio-Rad Laboratories, USA) at 150 volts with a constant pulse time of 14 sec for 16 h.

Screening of the Cara BAC library with marker IPM3 identified three BAC clones: BAC167, BAC191 and BAC117, with potato DNA inserts ranging from 100 to 120 kb (Fig. 1C). *Dde*I digestion of the IPM3 DNA in these BAC clones and other potato DNA samples revealed that BAC117 carried the IPM3 allele that was linked in *cis* to *Rx*1. The other two BAC clones, BAC167 and BAC191, contained alleles from a corresponding region of the *rx* chromosomes. To identify the relative genome positions of these BAC clones, pairs of PCR primers were designed based on the sequence of the right and left ends of the insert. Inverse polymerase chain reaction (IPCR; Ochman *et al.*, 1990) was used to isolate the right and left end sequences of insert DNAs. BAC DNA was isolated and digested separately with *Nla*III, *Hpa*II, *Mse*I, *Hin*P1I, *Pvu*II, *Hae*III (for isolation of a left end sequence) or with *Rsa*I, *Sac*I, *Eco*RI, *Hae*III, *Mae*II, *Mse*I, *Pvu*II, *Hin*P1I (for isolation of a right end sequence) for 4 h at 37°C and recircularised by self ligation for 2 h at 20°C. Ligations were carried out using 5-50 ng of digested DNA and 5-10 units of T4 DNA LIGASE (Boehringer Mannheim, Germany) in a total volume of 100 µl. PCR amplification of the recircularised DNA was carried out using 3 µl of self-ligated DNA as the template. AB1 (5'-CCTAAATAGCTTGGCGTAATCATG-3') and AB2 (5'-TGACACTATAGAATACTCAAGCTT-3') primers were used for PCR amplification of the left end sequence of insert DNA. AB3 (5'-CGACCTGCAGGCATGCAAGCTT-3') and AB4 (5'-ACTCTAGAGGATCCCCGGGTAC-3') primers were used for PCR amplification of the right end sequence of insert DNA. PCR conditions were as follows: 94°C for 15 sec, 60°C for 15 sec, 72°C for 90 sec - for 35 cycles. PCR products were digested simultaneously with *Hin*dIII and the restriction enzyme used in the preparation

of IPCR DNA template. The released insert ends were gel purified and cloned into pGEM-3Z(f+) (Promega, USA). Sequences of the clones containing ~1-2 kb inserts were determined using a 377 or 373 DNA SEQUENCING SYSTEM (Applied Biosystems, UK). PCR tests using the BAC DNAs as templates showed that the BAC clones identified with marker IPM3 overlapped in the order BAC167, BAC117, BAC191, *Rx*1 (Fig. 1C). The 191L marker was separated from *Rx*1 by only a single chromosomal recombination event (in plant S1-1146; Fig. 1B) in a mapping population of 1720 plants. In the same population, 117L and IPM3 markers were separated from *Rx*1 by two and three recombination events respectively whereas the GP34 marker, present in BAC167, was separated from *Rx*1 by thirteen recombinations (Fig. 1B). The BAC library did not contain additional BACs extending further towards *Rx*1 from the 191L marker.

Screening of the Cara BAC library with IPM4, which mapped at 0.06 cM from *Rx*1 on the side away from IPM3 (Bendahmane *et al.*, 1997), identified two clones: BAC73 and BAC111, with inserts of ~70 kb each (Fig. 1C). *Taq*I digestion of the IPM4 CAPS marker in these clones suggested that BAC111 was linked in *cis* to the *Rx*1 locus but that BAC73 carries DNA insert from the *rx* chromosome. To determine the relative genome position of BAC111 and BAC73 PCR tests were performed using end sequence primers of these BAC clones (Table 1). These tests suggested that BAC73 overlaps with BAC111 and that 73L and 111L represent opposite ends of this set of overlapping BACs. Both 73L and 111L co-segregated with IPM4. In the initial mapping population of 1720 individuals, these markers were separated from *Rx*1 by one recombination event (in individual S1-761; Fig. 1B) and it was not possible to determine directly which of these markers was physically closer to *Rx*1. Hence, to orientate these BACs relative to *Rx*1, the Cara BAC library was screened with CAPS markers 111L and 73L. The BAC library was also screened with the IPM5 CAPS marker which is on the same side of *Rx*1 as IPM4, but further from *Rx*1 (Bendahmane *et al.*, 1997). It was anticipated that BACs containing IPM5 would orientate the 111L and 73L markers relative to *Rx*1. These analyses identified BAC218, carrying an allele of IPM5 identified by *Pst*I digestion, as being linked in *cis* to *Rx*1 (Fig. 1C). The end sequences of BAC218 insert DNA were converted into the CAPS markers, 218L and 218R, and mapped genetically to the recombination events between GP34 and IPM5. Marker 218L was positioned 0.48 cM (recombination fraction:8/1720) from *Rx*1, between IPM5 and CT129. The 218R marker was positioned between IPM4 and IPM5, 0.30 cM (recombination fraction: 5/1720) from

*Rx*1. A single BAC pool #29 was also identified which contains three markers: 218R, 73L and 111R. CAPS analysis revealed that each of these markers in the BAC pool #29 is represented by the allele linked in *cis* to *Rx*1. Hence, it was concluded that BAC pool #29 contains a single BAC clone, BAC29, with DNA insert linked in *cis* to *Rx*1. Therefore, BAC29 provided a link between BAC218 and the IPM4 BAC contig and orientated the markers from the IPM4 contig in the following order: *Rx*1, 111L, IPM4, 73L (Fig. 1B).

By screening the BAC library with 111L allele-specific primers BAC221 was identified which carries an insert DNA of 40 kb and is linked in *cis* to *Rx*1. The left end sequence of BAC221 is located inside of BAC111 whereas the right end sequence of BAC221 extends further towards *Rx*1 (Fig. 1C). However the marker 221R co-segregated with IPM4 in the S1-Cara mapping population and was separated from *Rx*1 by the recombination event in plant S1-761 (Fig. 1B).

To extend the IPM4 contig further towards *Rx*1 the Cara BAC library was screened with 221R allele-specific primers which identified BAC45 which has an insert DNA of 40 kb and is linked in *cis* to *Rx*1. The right end sequence of BAC45 is located inside of BAC221, whereas the left end sequence of BAC45, 45L, extends further towards *Rx*1 (Fig. 1C). However, BAC45 does not contain *Rx*1 as the CAPS marker 45L is genetically separated from *Rx*1 by the recombination event in plant S1-761 (Fig. 1B). Additional PCR screening of the BAC library with the 45L marker failed to identify any new BAC clones therefore leaving a gap between the IPM3 and IPM4 BAC contigs (Fig. 1C).

Taking into account that disease resistance loci in plants are often highly complex with small families of resistance genes clustered within several dozen kilobases (Ellis *et al.*, 1995; Hulbert and Bennetzen, 1991; Jones *et al.*, 1994; Martin *et al.*, 1993; Witham *et al.*, 1994), a low stringency PCR screening assay was developed for the identification of duplicated sequences related to CAPS markers from the vicinity of *Rx*1 (IPM3-IPM5 interval). Pools of DNA from 20 resistant plants (R pool) and 20 susceptible plants (S pool) and the individual BAC clones from the IPM4 contig were used as templates for PCR amplifications. Primer annealing temperatures in PCR reactions was 5 to 10°C lower than in conditions originally developed for each CAPS primer pair (Table 1) so that amplification of related sequences, in addition to the original marker, could take place. The PCR products obtained with a number of tested

CAPS primer pairs were the same size as the products produced under high stringency conditions. However, digestion of these low stringency PCR products with either *Taq*I, *Alu*I or *Dde*I restriction enzymes revealed several new DNA fragments that were not identified previously. These included fragments that were nonpolymorphic as well as fragments polymorphic between the R and S pools. Digestion of the low stringency IPM4 products from the R pool with *Taq*I identified the original IPM4 locus (designated IPM4a) in BAC111. There were also new IPM4 restriction fragments that had not been detected previously. One of these fragments (IPM4b) was nonpolymorphic in the R and S pools. This fragment originated from BAC221 as the *Taq*I restriction fragment of similar size was also detectable after digestion of the IPM4b allele derived from this BAC (Fig. 1B). A second new DNA fragment was polymorphic between R and S pools and was not detected after digestion of either IMP4a or IPM4b alleles derived from BAC111 and BAC221, respectively. This fragment cosegregated with *Rx*1 in all the plants of the S1-Cara mapping population, including plant S1-761 and others with recombination events between GP34 and IPM5. This new IPM4 marker allele was designated IPM4c (see Fig. 1B).

Screening of the Cara BAC library with IPM4 primers using conditions for the detection of the IPM4c allele identified a new BAC clone, BAC77, with a DNA insert of approximately 50 kb (Fig. 1C). The end fragments of BAC77 DNA insert were cloned, sequenced and converted into the CAPS markers 77L and 77R. Marker 77L co-segregated with both IPM4-c and *Rx*1 whereas 77R was separated from *Rx*1 by one recombination event in the recombinant individual S1-761 (Fig. 1B; based on analysis of 1720 segregants).

TABLE 1: Primer sequences and thermal cycling conditions for CAPS markers in the *Gpa2-Rx* interval.

Marker	Primers	PCR conditions	Restriction enzyme
GP34	5'-CGTTGCTAGGTAAGCATGAAGAAG 5'-GTTATCGTTGATTTCTCGTTCCG	94°C 15s 62°C 15s 72°C 1 min 35 cycles	*Taq*I
IPM3	5'-AGTAGTTTCAGGCTAGTG 5'-CAACATCACTTGATCAGAC	94°C 15s 54°C 15s 72°C 1 min 35 cycles	*Dde*I
117L	5'-CCTAGCGTAGAGCGGTGTATCCA 5'-GTAGACATTTAATAATTCGTCG	94°C 15s 57°C 20s 72°C 2 min 35 cycles	*Rsa*I
191L	5'-ACAAATTGTATAATTATAGTGATACG 5'-CAAGACATTAATTAACCAAACAATGG	94°C 15s 50°C 15s 72°C 2 min 35 cycles	*Eco*RI
77L	5'-GCTTCTAAACTCTAAATTCAGATTC 5'-CATGTGCGGACTCGTTCTTTTGTAG	94°C 15s 64°C 15s 72°C 1 min 35 cycles	*Alu*I

Marker	primers	PCR conditions	Restriction enzyme
IPM4	5'-GTACTGGAGAGCTAGTAGTGATCA 5'-GAACACCTTAACTACACGCTGCAGG	94°C 15s 62°C 15s 72°C 2 min 35 cycles	*Taq*I
77R	5'-CTCGAGGGATTGAATCCAAATTAT 5'-GGAAGCAGAATACTCCTGACTACT	94°C 15s 66°C 15s 72°C 1 min 35 cycles	*Hae*III
45L	5'-GGAGTCAATGCAGGGTCTATGGA 5'-CTCATTTGACACTTCTCGAACACA	94°C 15s 62°C 15s 72°C 1 min 35 cycles	allele specific
221R	5'-GCTTACATTTGCTCGAAGAAGCCAC 5'-CCTTAATAATCAATAGATTCAACTCG	94°C 15s 60°C 15s 72°C 1 min 35 cycles	allele specific
111R	5'-CCACTGTGTAAGGGTCAACTATAGTC 5'-GAGATGAAGATTTTCTTGTCTGATGG	94°C 15s 65°C 15s 72°C 1 min 30s 35 cycles	allele specific
73L	5'-CATTTCCTGAATTGCTTCCGACTTC 5'-CCATGAAAATTGTTATCACTGAGGTC	94°C 15s 60°C 15s 72°C 1 min 35 cycles	*Alu*I
218R	5'-GATTACAGTTGTGAATTAGTTCGGTA 5'-GCAACAGATATATTCCACTTACCATTC	94°C 15s 62°C 15s 72°C 1 min 30s 35 cycles	*Alu*I

EXAMPLE 4: FINE MAPPING OF THE *Gpa2* LOCUS

Cosegregation of *Gpa2* and *Rx*1 resistance in both the mapping populations initially used to map the two loci, F1SHxRH and S1-Cara, respectively, delimited the *Gpa2* locus to the IPM3-IPM5 interval (see Example 2). For fine-mapping of the *Gpa*2 locus, a total of 2,788 S1-Cara genotypes were assayed for recombination events in the IPM3-IPM5 interval. In addition 598 F1SHxRH genotypes were subjected to a GP34/IPM5 marker screening as marker IPM3 is not informative in population F1SHxRH. The GP34 CAPS marker is derived from a sequenced insert of RFLP clone GP34. The CAPS marker screening provided a total of 20 recombinants in the S1-Cara population and 9 recombinants in the F1SHxRH population. These recombinants were subsequently tested for the presence of markers 77L, IPM4c, 77R, 45L, 221R, IPM4a, 111R, 73L and 218R, all of which are derived from the PM3-IPM5 interval (see Fig. 2B), as well as for *Gpa2* resistance. The *Gpa2* resistance test was carried out using *G. pallida* population D383 (Rouppe van der Voort *et al.* 1997a). The nematode resistance assays were performed on plants derived from *in vitro* stocks, stem cuttings or tubers. *In vitro* plants were transferred from MS medium containing 3% saccharose to a mixture of silversand and sandy loam under a moist chamber for one week. Two to four weeks after planting, plants showing vigorous growth were inoculated with nematodes. Assays were further performed as described for stem cuttings and tubers as described in Example 1 and in Rouppe van der Voort *et al.* (1997a). *G. pallida* Rookmaker with different virulence characteristics as *G. pallida* D383 (Bakker *et al.* 1992) was used to confirm the specificity of *Gpa2* resistance in tested plants.

This analysis showed that *Gpa2* is located between markers IPM4c and 111R (Fig. 2B). Among the 2,788 S1-Cara genotypes and 598 F1SHxRH genotypes tested, only one genotype, S1-761, was identified in which a recombination event had occurred between *Gpa2* and marker 77R. On the other side of *Gpa2*, genotype S1-B811 identified marker 111R as a flanking marker for the *Gpa2* interval.

Marker orders deduced from the analysis of F1SHxRH corresponded to those found in population S1-Cara. Estimates of recombination frequencies and their standard errors were calculated with the aid of the program Linkage-1 (Suiter et al. 1983) by choosing the appropriate genetic model for each cross. Data for the non-recombinant class of genotypes were set for either a 3:1 segregation ratio for population S1-Cara or a

1:1 segregation ratio for population F1SH×RH since only strongly skewed segregation ratios will influence estimates of recombination frequencies notably (Säll and Nilsson 1994; Manly 1994). A chi-square test was used to test for differences in recombination frequencies between the marker intervals. The chi-square test criterion was determined from the recombinant and non-recombinant classes for each marker interval. Differences (rejection of the null hypothesis) were significant when the test criterion was greater than the $X^2_{[0.05]}$ value. Estimates of recombination frequencies deduced from both populations were merged to obtain an estimate of the average recombination value for each marker interval. The graphical genotypes (Young and Tanksley, 1992) shown in Fig. 1 display the boundaries of the *Gpa2* interval.

EXAMPLE 5: CONSTRUCTION OF A CONTIGUOUS BAC CONTIG SPANNING THE *Gpa2* LOCUS

Example 3 describes the preparation of a Cara BAC library from a progeny of a selfed cv. Cara and the identification and isolation of BAC clones BAC77, BAC45, BAC221 and BAC111, which map to the 0.06 cM IPM4c-111R genetic interval harbouring the *Gpa2* locus (Fig. 1C). Additional PCR screening of the Cara BAC library with markers 45L and 77R failed to identify any BAC clones that spanned the region between BAC77 and BAC45.

To bridge this gap between BAC77 and the IPM4 BAC contig (see Fig. 2C), a second BAC library was constructed from the diploid potato genotype SH83-92-488. High molecular weight potato DNA was prepared in agarose plugs from potato nuclei as described in Liu *et al.* (1994) with the following modifications. Plant nuclei were isolated by grinding leaf tissue (10 g) in liquid nitrogen, suspending the powder in 100 ml nuclei isolation buffer (10 mM Tris-HCl pH 9.5, 10 mM EDTA, 100 mM KCl, 0.5 M sucrose, 4 mM spermidine 1.0 mM spermine, 0.1% mercaptoethanol) and sequential filtering through one layer each of 280, 88, 55 and 20 μm nylon mesh. One-twentieth volume of isolation buffer supplemented with 10% Triton X-100 was added to the filtrate and left on ice for 15 min. The nuclei were pelleted at 4°C (in 50 ml screwcap tubes) at 2200 rpm for 10 min and resuspended with isolation buffer to a final volume of 1 ml. The nuclei were heated at 42°C for 1-2 min, mixed gently with an equal volume of 1.4% low-melting point inCert agarose (FMC) prepared in 10 mM Tris-HCl pH 9.5

and 10 mM EDTA and immediately poured into molds to form plugs (V=100 µl/plug). The agarose plugs were treated with lysis buffer (1% sarkosyl, 0.4 M EDTA pH 8.5, 0.2 mg/ml proteinase K and 3.8 mg/ml sodiumdisulfite) at 50 °C for 2 days with one change of lysis buffer. Proteinase K activity was inhibited by incubating the agarose plugs 12 hours at 50°C in $T_{10}E_{10}$ buffer (10 mM Tris-HCl pH 8.0, 10 mM EDTA) supplemented with 40 µg/ml PMSF.

Restriction enzym digestion of the agarose plugs and subsequent size selection was carried out essentially as described in Example 3, with the following modifications. Half of each plug (~10 µg DNA) was digested with 10 U of *Hin*dIII restriction enzym for 1 h. Size selection was carried out in two steps. Partially digested *S. tuberosum* DNA was initially subjected to CHEF electrophoresis at 4°C in 0.5 X TBE using a linear increasing pulse time of 60-90 sec and a field strength of 6 V/cm for 18 hr. After electrophoresis, the lanes containing the lambda DNA ladder were removed and stained with ethidium bromide to locate the region of the gel containing potato DNA fragments ranging from 100 to 150 kb in size. This region was excised from the gel using a glass coverslip and subjected to a second size selection step in a 1% SeaPlaque (low-melting point) agarose gel (FMC). CHEF electrophoresis was carried out for 10 hr at 4°C using a field strength of 4 V/cm and a constant pulse time of 5 sec. The compression zone containing DNA fragments of 100 kb was excised from the gel as described above and dialysed against 50 ml TE for 2 hr at 4°C. Dialysed agarose slices were then transferred to a 1.5 ml Eppendorf tube, melted at 70°C for 5 min and digested with 1 unit of GELASE (Epicentre Technologies, USA) per 100 mg of agarose gel for 1 hr at 45°C.

Ligation of the size selected DNA to *Hin*dIII-digested and dephosphorylated pBeloBAC11 and subsequent transformation of ElectroMAX *E. coli* DH10B competent cells (Life Technologies, UK) with the ligated DNA was carried as described in Example 3, using the BioRad Gene Pulser for electroporation. Settings on the BioRad Gene Pulser were as recommended for *E. coli* by the manufacturer. Approximately 60.000 white colonies were picked individually into 384 well microtiter plates containing LB freezing buffer, grown at 37°C for 24 hr and stored at -80°C.

By screening the SH BAC library, as described in Example 3, with CAPS markers 77R and 45L BAC clone SHBAC43 was identified (see Fig. 2C). For further analysis of SHBAC43, plasmid DNA was isolated from 5 ml overnight cultures (LB

medium supplemented with 12.5 mg/ml chloramphenicol) using the standard alkaline lysis miniprep protocol (Engebrecht *et al.*, 1997) and resuspended in 50 µl TE. Plasmid DNA (10 µl) was digested with *Not*I for 3 h at 37°C to release the insert DNA from the pBeloBAC11 vector, and subsequently analysed by CHEF electrophoresis. Comparison of the electrophoretic mobility of the SHBAC43 insert with that of the lambda concatemer ladder (BioRad) lead to the conclusion that SHBAC43 contains a BAC insert of approximately 110 kb.

EXAMPLE 6: IDENTIFICATION OF CANDIDATE RESISTANCE GENE HOMOLOGUES (RGH) WITHIN THE *Gpa2* LOCUS

Identification of candidate RGHs

Screening of BAC clones SHBAC43, BAC45, BAC221a and BAC111 with degenerate primers RG1 and RG2 based on conserved motifs within the NBS of the cloned resistance genes RPS2, N and L6 (see ; Aarts *et al*, 1998) resulted in the weak amplification of a 530 bp fragment from BAC221a. The use of this fragment as a probe to screen a Southern blot containing *Eco*RI digested DNA of SHBAC43, BAC45, BAC221a and BAC111 showed that SHBAC43 contained 2 copies of this sequence and that BAC clones BAC221a and BAC111 each contained one copy of this sequence.

Sequence analysis

The DNA sequence of BAC clones SHBAC43, BAC221a and BAC111 was determined by shotgun sequence analysis. For each BAC clone a set of random subclones with an average insert size of 2 kb was generated. 10 µg of CsCl purified DNA was sheared for 6 seconds on ice at 6 amplitude microns in 200 µl TE using a MSE soniprep 150 sonicator. After ethanol precipitation and resuspension in 20 µl TE the ends of the DNA fragments were repaired by T4 DNA polymerase digestion at 11°C for 25 minutes in a 50 µl reaction mixture comprising 1x T4 DNA polymerase buffer (New England BioLabs, USA), 1mM DTT, 100 µm of all 4 dNTP's and 25 U T4 DNA polymerase (New England Biolabs, USA), followed by incubation at 65°C for 15 minutes. The sheared DNA was subsequently separated by electrophoresis on 1% SeaPlaque LMP agarose gel (FMC). The fraction with a size of 1.5-2.5 kb was excised from the gel and dialysed against 50 ml TE for 2 hr at 4°C. Dialysed agarose slices were then transferred

to a 1.5 ml Eppendorf tube, melted at 70°C for 5 min, digested with 1 unit of GELASE (Epicentre Technologies, USA) per 100 mg of agarose gel for 1 hr at 45°C, and the DNA was subsequentty precipitated. The 1.5-2.5 kb fragments were ligated at 16°C in a *Eco*RV restricted and dephosphorylated pBluescript SK^+ vector (Stratagene Inc.). The ligation mixture was subsequently used to transform ElectroMAX *E. coli* DH10B competent cells (Life Technologies, UK) by electroporation using the BioRad Gene Pulser. Settings on the BioRad Gene Pulser were as recommended for *E. coli* by the manufacturer. The cells were spread on Luria broth (LB) agar plates containing ampicillin (100 µg/ml), 5-bromo-4-chloro-3-indolyl-ß-D-galactoside (Xgal) (64 µg/ml) and isopropyl-1-thio-ß-D-galactoside (IPTG) (32 µg/ml). Plates were incubated at 37°C for 24 hours. Individual white colonies were grown in 96-well flat-bottom blocks (1.5 ml Terrific Broth medium containing 100 µg/ml ampicillin).

Plasmid DNA was isolated using the QIAprep 96 Turbo Miniprep system in conjunction with the BioRobot™ 9600 (QIAGEN) according to the manufacturers instructions. The ABI PRISM dye terminator cycle sequencing ready kit was used to perform sequencing reactions in a PTC-200 Peltier Thermal Cycler (MJ Research). The DNA sequences of the clones were determined using standard M13 forward and reverse primers. Sequence assembly was done using the 1994 version of the STADEN sequence analysis programme (Dear and Staden, 1991).

Analysis of the contiguous sequence of each BAC clone using the computer programme GENSCAN (Burge and Karlin, 1997) and BLASTX (Altschul *et al.,* 1990) identified a total of four NBS/LRR containing genes, two on SHBAC43, one on BAC221a and one on BAC111 (Fig. 2D). Three of these sequences were designated candidates for the *Gpa2* gene and selected for complementation analysis; RGH1 on BAC221a, RGH2 on BAC111 and RGH3 on SHBAC43 (Fig. 2D). The second NBS/LRR gene identified on SHBAC43 contained marker IPM4c and is therefore outside of the *Gpa2* interval (Fig. 2D).

EXAMPLE 7: TRANSFORMATION

For complementation analysis a 5.5 kb S*st*I-*Xba*I genomic fragment containing RGH3 from SHBAC43 and two *Xba*I-*Xba*I genomic fragments of approximately 11 kb and 10.3 kb containing RGH1 or RGH2 from BAC221a and BAC111, respectively, were

subcloned into the plant transformation vector pBINPLUS (Van Engelen *et al.*, 1995). These binary plasmids, designated pBINRGH1-3 were transferred to *Agrobacterium tumefaciens* strain AGL0 (Lazo *et al.*, 1991) by electroporation using the BioRad Gene Pulser. Settings on the BioRad Gene Pulser were as recommended for *A. tumefaciens* by the manufacturer. The cells were spread on Luria broth (LB) agar plates containing kanamycin (100 mg/L) and rifampicin (50 mg/L). Plates were incubated at 28°C for 48 hours. Small-scale cultures from selected colonies were grown in LB medium containing kanamycin (100 mg/l) and rifampicin (50 mg/l). Plasmid DNA was isolated as described previously and the integrity of the plasmids was verified by restriction analysis upon reisolation from *A. tumefaciens* and subsequent transformation to *E. coli*. *A tumefaciens* cultures harbouring a plasmid with the correct DNA pattern were used to transform a susceptible potato genotype.

Transformation of the susceptible potato genotype, clone V, was essentially performed as described by Visser (1991) and is described briefly below. Stem explants (1 cm long internodes) were prepared from 5 week old tissue culture plants and precultured for 24 hours (25°C, 16 hour light regime) in Petri dishes containing 2 sterile filter papers saturated with PACM (feeding layers: MS30 medium supplemented with 2 g/l caseinehydrolysate, 1 mg/l 2,4 D and 0.5 mg/l kinetine, pH 5.8) which were placed on R3B medium (MS30 medium supplemented with 2 mg/l NAA and 1 mg/l BAP, pH 5.8). The explants were then infected for 10 minutes with an overnight culture of *A. tumefaciens* strain AGL0 containing either pBINRGH1, pBINRGH2, pBINRGH3 or the pBINPLUS vector, blotted dry on sterile filter paper and cocultured for 48 hours on the original feeder layer plates. Culture conditions were as described above. Overnight *A. tumefaciens* cultures were pelleted and resuspended in liquid MS20 medium prior to infection. Following cocultivation, the explants were transferred to MS20 medium (pH 5.8) supplemented with 1 mg/l zeatin, 200 mg/l cefotaxime, 200 mg/l vancomycin and 100 mg/l kanamycin and cultured under the culture conditions described above. The explants were transferred to fresh medium every 3 weeks. Emerging shoots were isolated and transferred to glass jars with selective medium lacking zeatin for root induction. Only transgenic shoots were able to root on the kanamycin containing medium.

EXAMPLE 8: COMPLEMENTATION ANALYSIS

In vitro grown transgenic (R_0) plants were initially subjected to the *in vitro* resistance assay as described in Example 1B whereby sterilized second stage juveniles of *G. pallida* popultion D383 were used as inoculum. Three nodes from four independent primary transformants of the 4 different transformations were assayed; R_0(RGH1), R_0(RGH2) and R_0(RGH3) transgenic plants contain the candidate genes RGH1, RGH2 and RGH3, respectively, and R_0(BINPLUS) transgenic plants are without insert DNA and function as control plants. In addition, three nodes from 12 *in vitro* grown resistant and 12 *in vitro* grown susceptible progeny plants derived from the F1SHxRH mapping population (see Example 2) were included in the assay. The results are shown in Table 2. The development of nematode females on the roots of R_0(RGH1), R_0(RGH3) and R_0(BINPLUS) plants was similar to that observed on the roots of the susceptible control plants. In contrast, the R_0(RGH2) plants showed the same incompatible interaction with *G. pallida* population D383 as the resistant control plants. Three lines of evidence indicate that the 10.3 kb DNA fragment, which is integrated in the genome of R_0(RGH2) plants, harbours the *Gpa2* gene. First, the number of females able to develop on the roots of R_0(RGH2) plants was equivalent to the number of females able to develop on the roots of resistant control plants. Second, 90% of all the females that developed on these plants remained small and were transluscent. This stagnation of female development was also observed on the roots of the resistant control plants. And third, the change in sex ratio (male/female=0.9) which is characteristic for the *Gpa2* phenotype was also observed for the R_0(RGH2) plants.

TABLE 2 . Results of the complementation assay for *Gpa2* resistance.

Genotype	Average no. cysts/3 plants (# genotypes)[1)]	Cyst phenotype
Susceptible F1SHxRH (*gpa2*)	42 (12)	White
Resistant F1SHxRH (*Gpa2*)	5 (12)	Transluscent
R_0(BINPLUS)	33 (4)	White
R_0(RGH1)	39 (4)	White
R_0(RGH2)	2 (4)	Transluscent
R_0(RGH3)	40 (4)	White

[1)] The numbers between brackets indicate the numbers of genotypes tested

Molecular and computer analysis of the genomic insert conferring resistance

To confirm the presence of the RGH2 insert in the R_0(RGH2) with the resistant phenotype, a marker analysis with CAPS marker IPM4 was performed. The presence of the RGH2 linked CAPS marker IPM4a in all the R_0(RGH2) plants transformed with pBINRGH2 indicates that the RGH2 gene is present in all these transgenic plants. Correct integration of the genomic fragment was also confirmed by Southern analysis using RGH2 and NPTII specific probes.

The sequence of the 10.3 kb *Xba*I-*Xba*I insert of pBINRGH2 is provided in Fig. 3C (SEQ ID NO.3). The genomic structure of the *Gpa2* gene was determined by analysis of the genomic sequence, derived from the insert of pBINRGH2, and cDNA fragments generated by 5' and 3' rapid amplification of cDNA ends (RACE). Comparison of these sequences revealed that *Gpa2* contains two introns at the C-terminus of the gene (Fig. 3). The first intron (237 bp) is located within the coding region of the gene whereas the second intron (112 bp) is situated within the 3'-untranslated region (Fig. 3). The second exon of *Gpa2* encodes a TGA stop codon and contains only 25 bp of coding nucleotides.

The deduced open reading frame (ORF) of the *Gpa2* gene encodes a predicted polypeptide of 912 amino acids with a MW of 104.5 KDa (SEQ ID NO.1).

EXAMPLE 9: IDENTIFICATION AND MAPPING OF HOMOLOGOUS GENES.

Screening of the SH83 BAC library as described in Example 4 using primers described in Leister *et al.* (1996) based on conserved motifs within the nucleotide binding site (NBS) of the cloned resistance gene RPS2 (GGVGKTT in case of primer S1 and GGLPLAL in case of primer AS1; see Tables 3 and 4) resulted in the amplification of DNA fragments of the expected sizes from all 156 BAC pools. This indicates that sequences homologous to the resistance gene motifs used to design primers S1 and AS1 are abundantly present in the potato genome.

Based on the nucleotide sequence of the resistance gene homologues RGH1-4, primers were designed for specific amplification of nucleic acid sequences cognate to the NBS of RGH1-4 (primers RG3 and RG4; see Tables 3 and 4). The position of primer RG3 corresponds to nucleotides 514-533 of SEQ ID NO.1 (Fig. 3). Primer RG4 is complementary to nucleotides 985-1002 of SEQ ID NO.1 (Fig. 3). These primers differ from RG1 and RG2 and those designed by Leister *et al.* (1996) in that the 3' terminal nucleotides are designed on the basis of amino acid residues that exceed the conserved residues used for the design of the former primers (see Table 4). PCR using primers RG3 and RG4 on template DNA of the BAC clones SHBAC43, BAC45, BAC221a and BAC111 resulted in amplification products of the expected size from SHBAC43, BAC221a and BAC111.

Screening of the SH83 BAC library as described in Example 4 using primers RG3 and RG4 identified 19 individual BAC clones that showed amplification of DNA fragments of the expected size. This indicates that these primers discriminate between RGH1-4 homologues and sequences containing common resistance gene motifs.

Primer sequences RG5 and RG6 (see Table 3) were designed on the basis of sequences outside of the NBS of RGH1-4. The position of primer RG5 corresponds to nucleotides 199-221 of SEQ ID NO.2 (Fig. 3). Primer RG6 is complementary to nucleotides 2681-2701 of SEQ ID NO.2 (Fig. 3). Screening the SH83 BAC library as described in Example 4 resulted in the isolation of 5 BAC clones which already were identified with primers RG3 and RG4. These BAC clones showed overlap with clones SHBAC43, BAC221a and BAC111. The primers RG5 and RG6 therefore discriminate between RGH sequences derived from the *Gpa2* locus and homologous variants

elsewhere on the potato genome. Primers RG3, 4, 5, 6 are SEQ ID NO. 4, 5, 6 and 7 respectively.

Mapping of the *Gpa2* homologues identified with primers RG3 and RG4 is carried out by developing CAPS markers designed on the end sequences of each BAC insert. These CAPS markers are used to screen 136 genotypes of population F1SHxRH. The data on marker segregation are scored and the respective loci are mapped on the SH83 genome by use of the computer package JoinMap2.0 (Stam, 1993). It is likely that one or more of these homologues map to regions of the potato genome harbouring mono- or polygenic resistance loci that confer resistance to other *G. pallida* or *G. rostochiensis* populations; such as *H1* (Pineda *et al.* 1993; Gebhardt *et al.* 1993), *Gpa* (Kreike *et al.* 1994), *Gpa5* (Rouppe van der Voort and van der Vossen; unpublished data) and *Grp1* (Rouppe van der Voort *et al.* 1998) on chromosome 5; *Gro1* on chromosome 7 (Barone *et al.*, 1990; Ballvora *et al.*, 1995); *Gpa6* on chromosome 9 (Rouppe van der Voort and van der Vossen; unpublished data) and *Gpa3* on chromosome 11 (P. Wolters, unpublished data).

Table 3: Primer sequences and thermal cycling conditions for identification of *Gpa2* homologues

Primer	Primer sequence[1)]	PCR conditions	Expected product size
s1 as1	5'-GGTGGGGTTGGGAAGACAACG 5'-TGCTAGAGGTAATCCTCC	94°C 30s 51°C 30s 72°C 2 min 35 cycles	500 bp
RG1 RG2	5'-GGIATGGGIGGIGTIGGIAARACNACN 5'-ICCIAGIACYTTIARIGCIARIGGIARWCC	94°C 30s 50°C 30s 72°C 2 min 30 cycles	530 bp
RG3 RG4	5'-GGAGGCATCGGGAAAACAAC 5'-TGCTAGAGGTAACCCTCC	94°C 30s 55°C 30s 72°C 2 min 30 cycles	488 bp
RG5 RG6	5'-GATATGGTTGACTCGGAATCAAG 5'-GAGTATGGACCTCGATAGAGC	94°C 30s 60°C 30s 72°C 3 min 30 cycles	2500 bp

[1)] R=A or G; Y=T or C; W=A or T

TABLE 4. Oligonucleotides based on conserved peptide motifs within the NBS of PPS2 and RGHs

Motif / primer	Primer designation	Sequence[2)]
P-loop (RPS2/N/L6) s1	sense	G G V G K T T ggt ggg gtt ggg aag aca acg
P-loop (RGH1-4) RG3	sense	G G I G K T T gga ggc atc ggg aaa aca ac
GLPLAL (RPS2/N/L6) as1	antisense[1)]	G L P L A L caa cgc tag tgg caa tcc
GGLPLA (RGH1-4) RG4	antisense[1)]	G G L P L A tgc tag agg taa ccc tcc

1) Antisense primers are written in opposite orientation to the peptide sequence
2) Differences between primers s1/as1 and primers RG3/RG4 are underlined

EXAMPLE 10: A MARKER ASSISTED SELECTION ASSAY FOR *Gpa2*

The *Gpa2* locus is hypothesized to be introgressed from *S. tuberosum* spp. *andigena* CPC1673 into European cultivars. Flanking markers tightly linked to *Gpa2* are likely to be diagnostic for the presence of *Gpa2* in these cultivars. Therefore, *Gpa2* linked CAPS markers were used to screen two clones (abbreviated as CPC1673-a and CPC1673-b) of the wild species *Solanum tuberosum* spp. *andigena* CPC 1673 (hereafter referred to as CPC1673) as well as nine potato cultivars harbouring introgressions from CPC1673. The CAPS marker profiles were highly similar for the selfed CPC1673 genotypes and the analyzed potato cultivars harboring introgressions from CPC1673. The CAPS marker alleles linked to *Gpa2* were only identified in regions which appeared to be of CPC1673 origin. Among the seven CPC1673 cultivars tested, five differences in the size of an

introgressed region of 0.9 cM were observed. All *Gpa2* containing cultivars harbored the *Gpa2* flanking markers 77R and 111R thereby demonstrating that these markers are indicative for the presence *Gpa2* (see Table 5).

TABLE 5: Potato clones having *S. tuberosum* spp. *andigena* CPC1673 in their pedigree (with the exception of clone RH89) tested on the presence of chromosome *12* specific CAPS alleles. Resistance or susceptibility to *G. pallida* population Pa2-D383 is indicated by "R" or "S" respectively. Presence or absence of a CAPS marker band that cosegregates with resistance in populations S1-Cara and F1SHxRH is indicated by either a "+" or a "-". The order of the presented CAPS markers corresponds to the marker order on chromosome *12*.

Clone	*Gpa2*	IPM3	191L	77L	IPM4c	77R	IPM4	111R	73L	218R	IPM5
CPC1673-a	n.d.	+	+	+	+	+	+	+	+	+	+
CPC1673-b	n.d.	+	+	+	+	+	+	+	+	+	+
Cara	R[a)]	+	+	+	+	+	+	+	+	+	+
Alcmaria	R[b)]	-	+	+	+	+	+	+	+	+	+
Multa	R[a)]	-	-	+	+	+	+	+	+	+	+
SH83	R[a)]	-	-	-	+	+	+	+	+	+	+
Amaryl	R[b)]	-	-	-	+	+	+	+	+	+	+
Marijke	R[b)]	-	-	-	+	+	+	+	+	+	+
Saturna	S[a)]	-	-	-	-	-	-	-	-	-	+
RH89	S[a)]	-	-	-	-	-	-	-	-	-	-

[a)] As determined by cyst counts on at least three replicates

[b)] Data from Arntzen et al. (1994)

FIGURES

Fig. 1. High resolution map of the *Rx* locus (not drawn to scale). **A.** Simplified genetic map of potato chromosome XII (denoted by a horizontal line) in which the area left of the arrow is reversed between the potato and tomato genetic maps (Tanksley *et al.*, 1992). Vertical lines indicate positions of previously mapped RFLP markers (Bendahmane *et al.*, 1997; Tanksley *et al.*, 1992). The filled rectangle denotes a genetic interval between markers GP34 and 218L, which is magnified in panels B and C. **B.** Genetic map of the GP34-218L interval (denoted by a horizontal line). Positions of the RFLP marker GP34 and the AFLP markers IPM3, IPM4a and IPM5 are indicated by vertical lines. The positions of BAC end-derived markers and low-stringency PCR markers (enclosed in square brackets) are indicated by vertical arrows. The symbols L and R denote the BAC ends that were mapped relative to Rx1. The numbers in brackets below the bar indicate the numbers of S1-Cara individuals containing recombination events in each marker interval, identified in the initial S1-Cara mapping population of 1720 individuals. The predicted position of *Rx*1, delimited by the cross-over events in plants S1-1146 and S1-761, is indicated by the horizontal arrow. **C.** Positions of Cara BAC clones in the GP34-218L interval. Each open rectangle represents one BAC insert DNA. Inside of each rectangle is the name of the BAC clone, the estimated insert size in kb (except for the BAC29).

Fig. 2. High resolution genetic and physical map of the Gpa2 locus. **A.** Relative position of the *Gpa2* locus on chromosome 12 of potato. Vertical lines indicate positions of previously mapped RFLP markers. The filled rectangle denotes the Gpa2 locus between markers IPM3 and IPM5 which is magnified in panel B. **B.** High resolution genetic map and graphical genotypes of the IPM3-IPM5 interval, indicating differences in the size of *Solanum tuberosum* spp. *andigena* CPC1673 derived segments in different potato genotypes. The relative positions of CAPS markers used to fine-map *Gpa2* are indicated by vertical bars. The presented genotypes border the *Gpa2* interval. Introgression segments are indicated by thick bars. Size of marker intervals are not drawn to scale. The symbols R (for resistant) and S (for susceptible) indicate the *Gpa2* phenotype of the tested genotypes. **C.** High resolution physical map

of the *Gpa2* locus. The relative positions of CAPS markers are indicated by vertical bars. The open rectangles represent BAC clones isolated from the Cara BAC library. The shaded rectangle represents a BAC clone isolated from the SH83 BAC library. The name of each BAC clone is depicted within the rectangle and the estimated insert size is in given in kb. The predicted position of *Gpa2* is indicated by the horizontal arrow. Recombinant S1-Cara genotypes S1-761 and S1-B811 delimit the *Gpa2* genetic interval. **D.** Relative positions of four resistance gene homologues (RGH1-4) identified within the IPM4c-111R *Gpa2* interval.

Fig. 3. Nucleic and amino acid sequence of the *Gpa2* gene. **A.** Coding nucleic acid sequence and deduced amino acid sequence of the *Gpa2* gene. **B.** Coding sequence of the *Gpa2* gene including intron 1. The position of intron 1 is indicated in bold italics (position 2712-2948). **C.** Sequence of the 10.3 kb *Xba*I-*Xba*I genomic DNA fragment inserted in pBINRGH2, harbouring the *Gpa2* gene. The initiation ATG codon (position 4875-4877) and the termination TGA codon (position 7848-7850) are underlined. The positions of intron 1 (7586-7822) and intron 2 (7942-8053) are indicated in bold italics.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: CPRO-DLO
(B) STREET: Droevendaalsesteeg 1
(C) CITY: Wageningen
(D) STATE: Gelderland
(E) COUNTRY: The Netherlands
(F) POSTAL CODE (ZIP): Postbus 16 6700 AA

(A) NAME: Landbouw Universiteit Wageningen
(B) STREET: Dreyenlaan 2
(C) CITY: Wageningen
(D) STATE: Gelderland
(E) COUNTRY: Netherlands
(F) POSTAL CODE (ZIP): Postbus 9101 6700 HB

(ii) TITLE OF INVENTION: Engineering nematode resistance in Solanacae

(iii) NUMBER OF SEQUENCES: 7

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2739 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: *Gpa2* encoding sequence of S. tuberosum

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..2739

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
atg gct tat gct gct gtt act tcc ctt atg aga acc ata cat caa tca   48
Met Ala Tyr Ala Ala Val Thr Ser Leu Met Arg Thr Ile His Gln Ser
  1               5                  10                  15

atg gaa ctt act gga tgt gat ttg caa ccg ttt tat gaa aag ctc aaa   96
Met Glu Leu Thr Gly Cys Asp Leu Gln Pro Phe Tyr Glu Lys Leu Lys
            20                  25                  30

tct ttg aga gct att ctg gag aaa tcc tgc aat ata atg ggc gat cat  144
Ser Leu Arg Ala Ile Leu Glu Lys Ser Cys Asn Ile Met Gly Asp His
         35                  40                  45

gag ggg tta aca atc ttg gaa gtt gaa atc ata gag gta gca tac aca  192
Glu Gly Leu Thr Ile Leu Glu Val Glu Ile Ile Glu Val Ala Tyr Thr
     50                  55                  60

aca gaa gat atg gtt gac tcg gaa tca aga aat gtt ttt tta gca cgg  240
Thr Glu Asp Met Val Asp Ser Glu Ser Arg Asn Val Phe Leu Ala Arg
 65                  70                  75                  80

aat gtg ggg aaa aga agc agg gct atg tgg ggg att ttt ttc gtc ttg  288
Asn Val Gly Lys Arg Ser Arg Ala Met Trp Gly Ile Phe Phe Val Leu
                85                  90                  95

gaa caa gca cta gaa tgc att gat tcc acc gtg aaa cag tgg atg gca  336
Glu Gln Ala Leu Glu Cys Ile Asp Ser Thr Val Lys Gln Trp Met Ala
            100                 105                 110

aca tcg gac agc atg aaa gat cta aaa cca caa act agc tca ctt gtc  384
Thr Ser Asp Ser Met Lys Asp Leu Lys Pro Gln Thr Ser Ser Leu Val
        115                 120                 125

agt tta cct gaa cat gat gtt gag cag ccc gag aat ata atg gtt ggc  432
Ser Leu Pro Glu His Asp Val Glu Gln Pro Glu Asn Ile Met Val Gly
    130                 135                 140

cgt gaa aat gaa ttt gag atg atg ctg gat caa ctt gct aga gga gga  480
Arg Glu Asn Glu Phe Glu Met Met Leu Asp Gln Leu Ala Arg Gly Gly
145                 150                 155                 160

agg gaa cta gaa gtt gtc tca atc gta ggg atg gga ggc atc ggg aaa  528
Arg Glu Leu Glu Val Val Ser Ile Val Gly Met Gly Gly Ile Gly Lys
                165                 170                 175

aca act ttg gct gca aaa ctc tat agt gat cct tac att atg tct cga  576
Thr Thr Leu Ala Ala Lys Leu Tyr Ser Asp Pro Tyr Ile Met Ser Arg
            180                 185                 190

ttt gat att cgt gca aaa gca act gtt tca caa gag tat tgt gtg aga  624
Phe Asp Ile Arg Ala Lys Ala Thr Val Ser Gln Glu Tyr Cys Val Arg
        195                 200                 205

aat gta ctc cta ggc ctt ctt tct ttg aca agt gat gaa cct gat tat  672
Asn Val Leu Leu Gly Leu Leu Ser Leu Thr Ser Asp Glu Pro Asp Tyr
    210                 215                 220
```

```
cag cta gcg gac caa ctg caa aag cat ctg aaa ggc agg aga tac ttg   720
Gln Leu Ala Asp Gln Leu Gln Lys His Leu Lys Gly Arg Arg Tyr Leu
225                 230                 235                 240

gta gtc att gat gac ata tgg act aca gaa gct tgg gat gat ata aaa   768
Val Val Ile Asp Asp Ile Trp Thr Thr Glu Ala Trp Asp Asp Ile Lys
                245                 250                 255

cta tgt ttc cca gac tgc gat aat gga agc aga ata ctc ctg act act   816
Leu Cys Phe Pro Asp Cys Asp Asn Gly Ser Arg Ile Leu Leu Thr Thr
            260                 265                 270

cgg aat gtg gaa gtg gct gaa tat gct agc tca ggt aag cct cct cat   864
Arg Asn Val Glu Val Ala Glu Tyr Ala Ser Ser Gly Lys Pro Pro His
        275                 280                 285

cac atg cgc ctc atg aat ttt gac gaa agt tgg aat tta cta cac aaa   912
His Met Arg Leu Met Asn Phe Asp Glu Ser Trp Asn Leu Leu His Lys
    290                 295                 300

aag atc ttt gaa aaa gaa ggt tct tat tct cct gaa ttt gaa aat att   960
Lys Ile Phe Glu Lys Glu Gly Ser Tyr Ser Pro Glu Phe Glu Asn Ile
305                 310                 315                 320

ggg aaa caa att gca tta aaa tgt gga ggg tta cct cta gca att act   1008
Gly Lys Gln Ile Ala Leu Lys Cys Gly Gly Leu Pro Leu Ala Ile Thr
                325                 330                 335

ttg att gct gga ctt ctc tcc aaa atc agt aaa aca ttg gat gag tgg   1056
Leu Ile Ala Gly Leu Leu Ser Lys Ile Ser Lys Thr Leu Asp Glu Trp
            340                 345                 350

caa aat gtt gcg gag aat gta cgt tcg gtg gta agc aca gat ctt gaa   1104
Gln Asn Val Ala Glu Asn Val Arg Ser Val Val Ser Thr Asp Leu Glu
        355                 360                 365

gca aaa tgc atg aga gtg ttg gct ttg agt tac cat cac ttg cct tct   1152
Ala Lys Cys Met Arg Val Leu Ala Leu Ser Tyr His His Leu Pro Ser
    370                 375                 380

cac cta aaa ccg tgt ttt ctg tat ttt gca att ttc gca gag gat gaa   1200
His Leu Lys Pro Cys Phe Leu Tyr Phe Ala Ile Phe Ala Glu Asp Glu
385                 390                 395                 400

cgg att tat gta aat aaa ctt gtt gag tta tgg gcc gta gag ggg ttt   1248
Arg Ile Tyr Val Asn Lys Leu Val Glu Leu Trp Ala Val Glu Gly Phe
                405                 410                 415

ttg aat gaa gaa gag gga aaa agc ata gaa gag gtg gca gaa aca tgt   1296
Leu Asn Glu Glu Glu Gly Lys Ser Ile Glu Glu Val Ala Glu Thr Cys
            420                 425                 430

ata aac gaa ctt gta gat aga agt cta att tct atc cac aat gtg agt   1344
Ile Asn Glu Leu Val Asp Arg Ser Leu Ile Ser Ile His Asn Val Ser
        435                 440                 445
```

```
ttt gat ggg gaa aca cag aga tgt gga atg cat gat gtg acc cgt gaa   1392
Phe Asp Gly Glu Thr Gln Arg Cys Gly Met His Asp Val Thr Arg Glu
    450                 455                 460

ctc tgt ttg agg gaa gct cga aac atg aat ttt gtg aat gtt atc aga   1440
Leu Cys Leu Arg Glu Ala Arg Asn Met Asn Phe Val Asn Val Ile Arg
465                 470                 475                 480

gga aag agt gat caa aat tca tgt gca caa tcc atg cag tgt tcc ttt   1488
Gly Lys Ser Asp Gln Asn Ser Cys Ala Gln Ser Met Gln Cys Ser Phe
                485                 490                 495

aag agt cga agt cgg atc agt atc cat aat gag gaa gaa ttg gtt tgg   1536
Lys Ser Arg Ser Arg Ile Ser Ile His Asn Glu Glu Glu Leu Val Trp
            500                 505                 510

tgt cgt aac agc gag gct cat tct atc atc acg ttg tgt ata ttc aaa   1584
Cys Arg Asn Ser Glu Ala His Ser Ile Ile Thr Leu Cys Ile Phe Lys
        515                 520                 525

tgc gtc aca ctg gaa ttg tct ttc aag cta gta aga gta cta gat ctt   1632
Cys Val Thr Leu Glu Leu Ser Phe Lys Leu Val Arg Val Leu Asp Leu
    530                 535                 540

ggt ttg act aca tgc cca att ttt ccc agt gga gta ctt tct cta att   1680
Gly Leu Thr Thr Cys Pro Ile Phe Pro Ser Gly Val Leu Ser Leu Ile
545                 550                 555                 560

cat ttg aga tac cta tct ttg cgt ttt aat cct cgc tta cag cag tat   1728
His Leu Arg Tyr Leu Ser Leu Arg Phe Asn Pro Arg Leu Gln Gln Tyr
                565                 570                 575

cga gga tcg aaa gaa gct gtt ccc tca tca ata ata gac att cct cta   1776
Arg Gly Ser Lys Glu Ala Val Pro Ser Ser Ile Ile Asp Ile Pro Leu
            580                 585                 590

tcg ata tca agc cta tgc tat ctg caa act ttt aaa ctt tac cat cca   1824
Ser Ile Ser Ser Leu Cys Tyr Leu Gln Thr Phe Lys Leu Tyr His Pro
        595                 600                 605

ttt ccc aat tgt tat cct ttc ata tta cca tcg gaa att ttg aca atg   1872
Phe Pro Asn Cys Tyr Pro Phe Ile Leu Pro Ser Glu Ile Leu Thr Met
    610                 615                 620

cca caa ttg agg aag ctg tgt atg ggc tgg aat tac ttg cgg agt cat   1920
Pro Gln Leu Arg Lys Leu Cys Met Gly Trp Asn Tyr Leu Arg Ser His
625                 630                 635                 640

gag cct aca gag aac aga ttg gtt ttg aaa agt ttg caa tgc ctc aat   1968
Glu Pro Thr Glu Asn Arg Leu Val Leu Lys Ser Leu Gln Cys Leu Asn
                645                 650                 655

gaa ttg aat cct cgg tat tgt aca ggg tct ttt tta aga cta ttt ccc   2016
Glu Leu Asn Pro Arg Tyr Cys Thr Gly Ser Phe Leu Arg Leu Phe Pro
            660                 665                 670
```

```
aat tta aag aag ttg gaa gta ttt ggc gtc aaa gag gac ttt cgc aat   2064
Asn Leu Lys Lys Leu Glu Val Phe Gly Val Lys Glu Asp Phe Arg Asn
        675                 680                 685

cac aag gac ctg tat gat ttt cgc tac tta tat cag ctc gag aaa ttg   2112
His Lys Asp Leu Tyr Asp Phe Arg Tyr Leu Tyr Gln Leu Glu Lys Leu
    690                 695                 700

gca ttt agt act tat tat tca tct tct gct tgc ttt cta aaa aac act   2160
Ala Phe Ser Thr Tyr Tyr Ser Ser Ser Ala Cys Phe Leu Lys Asn Thr
705                 710                 715                 720

gca cct tta ggt tct act ccg caa gat cct ctg agg ttt cag atg gaa   2208
Ala Pro Leu Gly Ser Thr Pro Gln Asp Pro Leu Arg Phe Gln Met Glu
                725                 730                 735

aca ttg cac tta gag act cat tcc agg gca act gca cct cca act gat   2256
Thr Leu His Leu Glu Thr His Ser Arg Ala Thr Ala Pro Pro Thr Asp
            740                 745                 750

gtt cca act ttc ctc tta cct cct ccg gat tgt ttt cca caa aac ctt   2304
Val Pro Thr Phe Leu Leu Pro Pro Pro Asp Cys Phe Pro Gln Asn Leu
        755                 760                 765

aag agt tta act ttt agc gga gat ttc ttt ttg gca tgg aag gat ttg   2352
Lys Ser Leu Thr Phe Ser Gly Asp Phe Phe Leu Ala Trp Lys Asp Leu
    770                 775                 780

agc att gtt ggt aaa tta ccc aaa ctc gag gtc ctt caa cta tca cac   2400
Ser Ile Val Gly Lys Leu Pro Lys Leu Glu Val Leu Gln Leu Ser His
785                 790                 795                 800

aat gcc ttc aaa ggc gag gag tgg gaa gta gtt gag gaa ggg ttt cct   2448
Asn Ala Phe Lys Gly Glu Glu Trp Glu Val Val Glu Glu Gly Phe Pro
                805                 810                 815

cac ttg aag ttc ttg ttt ctg gat agc ata tac att cgg tac tgg aga   2496
His Leu Lys Phe Leu Phe Leu Asp Ser Ile Tyr Ile Arg Tyr Trp Arg
            820                 825                 830

gct agt agt gat cac ttt cca tac ctt gaa cga ctt ttt ctt agc gat   2544
Ala Ser Ser Asp His Phe Pro Tyr Leu Glu Arg Leu Phe Leu Ser Asp
        835                 840                 845

tgc ttt tat ttg gat tca atc cct cga gat ttt gca gat ata acc aca   2592
Cys Phe Tyr Leu Asp Ser Ile Pro Arg Asp Phe Ala Asp Ile Thr Thr
    850                 855                 860

cta gct ctt att gat ata ttt cgc tgc caa caa tct gtt ggg aat tcc   2640
Leu Ala Leu Ile Asp Ile Phe Arg Cys Gln Gln Ser Val Gly Asn Ser
865                 870                 875                 880

gcc aag caa att caa cag gac att caa gac aac tat gga agc tct atc   2688
Ala Lys Gln Ile Gln Gln Asp Ile Gln Asp Asn Tyr Gly Ser Ser Ile
                885                 890                 895
```

```
gag gtc cat act cgt tat ctt tat cga aat gga gca ttt ttg gta gtg   2736
Glu Val His Thr Arg Tyr Leu Tyr Arg Asn Gly Ala Phe Leu Val Val
            900                 905                 910

tga                                                              2739
 *
```

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2976 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: *Gpa2* coding and non coding sequence of S. tuberosum

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```
ATGGCTTATG CTGCTGTTAC TTCCCTTATG AGAACCATAC ATCAATCAAT GGAACTTACT      60

GGATGTGATT TGCAACCGTT TTATGAAAAG CTCAAATCTT TGAGAGCTAT TCTGGAGAAA     120

TCCTGCAATA TAATGGGCGA TCATGAGGGG TTAACAATCT TGGAAGTTGA AATCATAGAG     180

GTAGCATACA CAACAGAAGA TATGGTTGAC TCGGAATCAA GAAATGTTTT TTTAGCACGG     240

AATGTGGGGA AAAGAAGCAG GGCTATGTGG GGGATTTTTT TCGTCTTGGA ACAAGCACTA     300

GAATGCATTG ATTCCACCGT GAAACAGTGG ATGGCAACAT CGGACAGCAT GAAAGATCTA     360

AAACCACAAA CTAGCTCACT TGTCAGTTTA CCTGAACATG ATGTTGAGCA GCCCGAGAAT     420

ATAATGGTTG GCCGTGAAAA TGAATTTGAG ATGATGCTGG ATCAACTTGC TAGAGGAGGA     480

AGGGAACTAG AAGTTGTCTC AATCGTAGGG ATGGGAGGCA TCGGGAAAAC AACTTTGGCT     540

GCAAAACTCT ATAGTGATCC TTACATTATG TCTCGATTTG ATATTCGTGC AAAAGCAACT     600

GTTTCACAAG AGTATTGTGT GAGAAATGTA CTCCTAGGCC TTCTTTCTTT GACAAGTGAT     660

GAACCTGATT ATCAGCTAGC GGACCAACTG CAAAAGCATC TGAAAGGCAG GAGATACTTG     720

GTAGTCATTG ATGACATATG GACTACAGAA GCTTGGGATG ATATAAAACT ATGTTTCCCA     780

GACTGCGATA ATGGAAGCAG AATACTCCTG ACTACTCGGA ATGTGGAAGT GGCTGAATAT     840

GCTAGCTCAG GTAAGCCTCC TCATCACATG CGCCTCATGA ATTTTGACGA AAGTTGGAAT     900
```

TTACTACACA AAAAGATCTT TGAAAAAGAA GGTTCTTATT CTCCTGAATT TGAAAATATT 960

GGGAAACAAA TTGCATTAAA ATGTGGAGGG TTACCTCTAG CAATTACTTT GATTGCTGGA 1020

CTTCTCTCCA AAATCAGTAA AACATTGGAT GAGTGGCAAA ATGTTGCGGA GAATGTACGT 1080

TCGGTGGTAA GCACAGATCT TGAAGCAAAA TGCATGAGAG TGTTGGCTTT GAGTTACCAT 1140

CACTTGCCTT CTCACCTAAA ACCGTGTTTT CTGTATTTTG CAATTTTCGC AGAGGATGAA 1200

CGGATTTATG TAAATAAACT TGTTGAGTTA TGGGCCGTAG AGGGGTTTTT GAATGAAGAA 1260

GAGGGAAAAA GCATAGAAGA GGTGGCAGAA ACATGTATAA ACGAACTTGT AGATAGAAGT 1320

CTAATTTCTA TCCACAATGT GAGTTTTGAT GGGGAAACAC AGAGATGTGG AATGCATGAT 1380

GTGACCCGTG AACTCTGTTT GAGGGAAGCT CGAAACATGA ATTTTGTGAA TGTTATCAGA 1440

GGAAAGAGTG ATCAAAATTC ATGTGCACAA TCCATGCAGT GTTCCTTTAA GAGTCGAAGT 1500

CGGATCAGTA TCCATAATGA GGAAGAATTG GTTTGGTGTC GTAACAGCGA GGCTCATTCT 1560

ATCATCACGT TGTGTATATT CAAATGCGTC ACACTGGAAT TGTCTTTCAA GCTAGTAAGA 1620

GTACTAGATC TTGGTTTGAC TACATGCCCA ATTTTTCCCA GTGGAGTACT TTCTCTAATT 1680

CATTTGAGAT ACCTATCTTT GCGTTTTAAT CCTCGCTTAC AGCAGTATCG AGGATCGAAA 1740

GAAGCTGTTC CCTCATCAAT AATAGACATT CCTCTATCGA TATCAAGCCT ATGCTATCTG 1800

CAAACTTTTA AACTTTACCA TCCATTTCCC AATTGTTATC CTTTCATATT ACCATCGGAA 1860

ATTTTGACAA TGCCACAATT GAGGAAGCTG TGTATGGGCT GGAATTACTT GCGGAGTCAT 1920

GAGCCTACAG AGAACAGATT GGTTTTGAAA AGTTTGCAAT GCCTCAATGA ATTGAATCCT 1980

CGGTATTGTA CAGGGTCTTT TTTAAGACTA TTTCCCAATT TAAAGAAGTT GGAAGTATTT 2040

GGCGTCAAAG AGGACTTTCG CAATCACAAG GACCTGTATG ATTTTCGCTA CTTATATCAG 2100

CTCGAGAAAT TGGCATTTAG TACTTATTAT TCATCTTCTG CTTGCTTTCT AAAAAACACT 2160

GCACCTTTAG GTTCTACTCC GCAAGATCCT CTGAGGTTTC AGATGGAAAC ATTGCACTTA 2220

GAGACTCATT CCAGGGCAAC TGCACCTCCA ACTGATGTTC CAACTTTCCT CTTACCTCCT 2280

CCGGATTGTT TTCCACAAAA CCTTAAGAGT TTAACTTTTA GCGGAGATTT CTTTTTGGCA 2340

TGGAAGGATT TGAGCATTGT TGGTAAATTA CCCAAACTCG AGGTCCTTCA ACTATCACAC 2400

AATGCCTTCA AAGGCGAGGA GTGGGAAGTA GTTGAGGAAG GGTTTCCTCA CTTGAAGTTC 2460

TTGTTTCTGG ATAGCATATA CATTCGGTAC TGGAGAGCTA GTAGTGATCA CTTTCCATAC 2520

CTTGAACGAC TTTTTCTTAG CGATTGCTTT TATTTGGATT CAATCCCTCG AGATTTTGCA 2580

GATATAACCA CACTAGCTCT TATTGATATA TTTCGCTGCC AACAATCTGT TGGGAATTCC 2640

GCCAAGCAAA TTCAACAGGA CATTCAAGAC AACTATGGAA GCTCTATCGA GGTCCATACT 2700

CGTTATCTTT A***GTAAGACAT CTTCTTCCTT GATTTACAAC AATATTTAAC TCATCATCAT*** 2760

AGTAAACTCG ATAATAATCT GGATAATAGC TTTAGTAAGT CAAATTGCAC CAATTCAACA 2820

AAAGTTCTTG ATGCTGTCAT TGTGATTGAT TCGAATCCTT CCAATATTGT GTAACTTGTT 2880

ATACTTGCAT GTTCATTCTT GATTTTGGGA AGTGTAACAT TTCCATTTTT CATCTTGATT 2940

TTGGGAAGTC GAAATGGAGC ATTTTTGGTA GTGTGA 2976

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10329 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: *Xba*I-*Xba*I pBINRGH2 fragment containing *Gpa2* promoter, coding and non coding sequence of S. tuberosum

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

CTAGAGATTG GAATGGAGTG ATTCTTAGGG GTTTCTTTTT GAATTAATAT GAGGGTTAGT 60

ATTCAATCTT CAATTCGACA TTTTCTCATA ATTTCTTTAT CTGTTTATTT TTCCTATTCG 120

TAAATCTCTT GGGAAAAATT GGGGTTTTAT CGATTTGGAC TCCTTTTTGA TGAAAAAGGT 180

ATATTTACGA TCTTTATGTT ATGGGTAAAC TGATTTTAAC ATAAAATTAT TGATTCATCG 240

ATTATTTTTA TCATATTAAC CGCGTACAAT TTGGACTTTC CCGGTAAAGT TAAAGTATGA 300

TAAATTGAGA ATTTCAAGGT CGATCTTAGC TCCATTTTTG ATGAAATTTC ATATTTGAAC 360

TTATCTAAGC ATGGGTAAGA TGTTTTTCAA GAAATATTTC ATTTTCGAGT CGGGGTTTTG 420

GATTCGAATA TTTTAGGCTT CTTCAAGAAT GTAGATTTTT GTTTAAATTG AGTTTGTGAA 480

TTGATTTCAA CTCCATTTTC AAATTGGTTT TCACCATTAG CTTCCAAATA CTTTAAGGAT 540

CATTTTACAT CAAAAAATTC CAGATTTGGG TATCGTTTTC CGGTATGAGA CTTTTGGACC 600

GTTTTGCCCC TTTTCCCTAA ATTTCTTGAT TTTGGTGTCA TTGGACTCGA ATTGTGATTG 660

TGAATAATTG TTTGAATAGA TTATCGTGAT CCAGATTATA CTTGGAAAGG AAAGGCTCAA 720

GTCAAGTAAC TTTTGGAGTT CGTTTTAAGG CAAGTGGCTT CCAAACTTTG TAAAACTCTT 780

AGACTACGCA TGACTACTTT CCTAATTATG TTGGGGAGTA ATGGGGGATT GAGGATGGGT 840

TTTATTTGTT GATTGAAATT GTTGTAAATG AAAGATGGGG AATAAAACGA GCTAAATGTG 900

TTATGTGTGA CTTGAATTTG TTTGAATAAG TCATGTGATA ACTGATATTG AGGGATAGAA 960

GAGCATGAGC AGGCTATGAT TGATACAGAC ATTGATGTTG AGGCAGATGA TGTGTAATAC 1020

TATGATGTGG TCGTGATATG GTTGTGATTG AGACATGTGA TGTGTAATAC TATGATGTGG 1080

TCGTGATATG GTTGTGATTG AGACAGGTGA TGTGTAATAC TATGATGTGG TCGTGATATG 1140

GTTGTGATTG AGACAGGTGA TGTGTAATAC TATGATGTGG TCGTGATATG GTTGTGACTG 1200

AGACAGGTGA TGTGTAATAC TATGATGTGG TCGTGATATG GTTGTGATTG AGACAGATGA 1260

TGTGTAATAC GATGATGTGA TCGTGATATG ATTGTGATTG ATTACATGTG CATATTCATT 1320

ATTCATCCCA TGTGTGAACT ATCTGTTGCA TGAGTTCTGA GACACTGATA TGAGGATGGA 1380

TGGATATGAG ACACAGTTGA GACTAGCTCC GGCTAGAGAT GTATGAGATG GACTAGCTCC 1440

GGCTAGCGAT TTGGATGCCG ATGGGATCTG GTTCCGGCGG TGATACATGG TCCATGTGTG 1500

GCCCCCATGG GTTCTGATTT GAGTATTCAA CGCGGACTGA TTACGTCAAC AGATGTGTAT 1560

CGTAGGACAG ACATGTATCA CGACTACATG ACATCATTAT TGCATTTTGC ATCGCATTTG 1620

CCTTATCTTT GTCTGTGATG TGTGGATTGT ATCGGTTTAC CCTTTTTATG TGGAATTTGA 1680

TCTACTTGCT CTTATTTGTT GATCTGAGGT TGATGAGGAT ATACTGTTGG TTCTGGCTGT 1740

TGAATATGAT CTGTTTAGTA TAGGTTGGTT GGTTTGCTGC TAGATTGAAG TTTCGGTGGT 1800

TCGGTTGGGA TTGAAAGGAG TTGTTTGTAG CTGCTAGTTT TGCTTAGTTT AGAGTTACTT 1860

GCGAGTACCT GTGGTTTTCG GTACTCACCC TTGCTTCTAC ACAATTGTGT AGGTTGACAG 1920

CTCTCTCTCA GATATTTTCT TTAGCAGATT GAGCTTTGAG ACATACTCGA GAGGTAGCGG 1980

TTCATTCCAG ACGTGCCCTT GAGTTATCTT TACTTTCAGT TTTGTTCTAT TCGAGAACTA 2040

TACTCTGAGA CTTGTATATT TTTATTCGAA TTCTGTATTT AGAGGTTTGT ACATGTGACA 2100

ACCAAATTCT GGGTAGTGTT AAGTCTTAAT TAAAGTTTTC TGCTTATTTA TTATCTTTTA 2160

TTCTCGTATT TCTACTTCTC TATCGTTGTG GTTGGGTTAG GCTGACGTGT CTGGTGGGAA 2220

ACGGACATGT GCCATCACAT CCGGATTTGG GGTGTGACAA ATATTTTGTT AGTTATATAC 2280

AAAATTGTAT GTAGTATATG TATATTTTCT GCTTTCATCA CAATTGTATA TAGATATTTG 2340

TATATTTTGT TAGTTATATA CAAAATTGCT TGAAGTATAT GTATATTTTC TGCTTAAATC 2400

ATAATTGTAT ATATATATAT ATATATATAT ATTTCTATAT TTTGTAAGTT ATATACAATA 2460

GTATGAATTA AACAATATAC AAACCTTACA TTATTATATA TACAGTTAGG TTACACCAAA 2520

AATTATCAAA TTAAAGCACA ACTTTTTTAT CGAATCATAT ACAATTCATA TATATAATTG 2580

ACTTAGTAAT TTTATACAAC TACTTACACT TCTACATGGT ATAAGAATTT TGCACAATTA 2640

CTTACATATA TACAATATTA TCAATTAAAC AATATACAAA TCGTATAACT TATATATACA 2700

GTAAAATTAC AACAACAACA ACAAAAATTA TCAAATTAAA GCACACCGTT GTTGTCGAAT 2760

CATATACACT CCATATATAC AAATTGTGTC ATTCAATTTT TCGAACAAAA AATTAGAATT 2820

GAATTGTTAA TATAAAATTT ATCTAATATT GTATAAACAA AATTAAATTA TTGCAAACCA 2880

TTAGAATGAA AAAAACAAAA ATAAACCGTT TTCCAAAATT TCAATTATAT ACTATACAAA 2940

TCAATTGTAT ACTTTCTTGC CGTTCAAAAC ATGAAGTTTC CTTGAAAGAA ACGCTTACCT 3000

AGCGTTGAAT ATACAAGAAT ATTGATTAAT CGTATGCTTC AGTCGTTTGA GGAACCCAGT 3060

TGTTATTGTG TTTCTATTGC TATAGAACTC CTTTTTGGAA AAATATTTGA TTTTGGACGA 3120

TTAGCTTGAA TCATGGGATT ATATAAAATT TTTATTACCG TATTTAGCAC TCATGTATCC 3180

ATTTATTAAA AAAAAATTGT ATAAATTATA TTTTTAAAAG AAAATATACA AAATTAATGC 3240

TTCATAGCAA ACTAAACTAT ACCCATTGAA TGTAATTACT AAACTATACC TATAGAGCGT 3300

TATTTCATTA AATACGTTTA TCATATATGA AGTTTTCCCT CAAGAGATCC TACACCTTAT 3360

ATATAGCTTC TCAAATGTGG AAATTCAATC TCACACCCAA CAATCTTTCC CTCAGACTAA 3420

GTTTCATGGC CCAATATCAC AATGATCCAC GAGTCAATTC ATGAGATTCA CTATGTGTGT 3480

CACCCACATC GTCTAAGTAT TTTATGGCAA TCAAGCCCTA CAACTTGCTT CTTCTTTATA 3540

TATATATATA TATATATATA TATATATATA TATATATGTG TGTGTGTGTG TGTGTGTGTG 3600

CGCATCTCTA ATTAATCTCG TAAAGGGATT AAGGGGCCAA TTTCAAAGAA TTAGGCGATT 3660

TTCTTAGTTT TTCGTGTGTG TTAACCCATA GGTATTTTGG TGATATGGTT TTCGGATGAT 3720

TTATTTTGTG CAACTTATAT GGAACCCTTC GTAGGGAGTT AGTCTCACAC TTTTTAGAGT 3780

CCATTTTGGG CATTCAGGGG CTAATTTATA GGAAATAGGT GATCTTCTCA GTTTGTCTGT 3840

ATTAGCCCAT GAATATTTTG GTGATATGTC TTCCGAATAA TTTCTTTGTA AAATCTTTAC 3900

GGGACCCTCC ATAGGGAGTT AGTGGAGCAG TACGTATAGT CTCACAATTT TAGAGTTCAT 3960

TTTGGGCATT TAGGGGCCAA TTTACAGGAT TTAGGCGACT TTCTCAGTGT TTTGTGTGTG 4020

TTAGCCCATT AATAGTTGGT GATATGACTT TCAGACGATT TCTTTGCTAC ACATTTACGG 4080

AACCCTCTGT AGGAAGTCGG GGGAGCAATA CGTACAATCT CACAATTTTA GAGTCCATTT 4140

TAGGCATTTA GGGGCCAATT TAAAGAAATT GGACAATTTT CTCAGTTTTT CGTGTCTGTT 4200

AGCCATTAAT ATATTGGTGA ATATGACCTA CAGATGATTT CTAATCGAAA TCTTTACGAA 4260

ACCCTCAGTA GGGAGTTGGG GGAGCAATAC GTACCGTCTG ACAATTTTTA GAGTCCATTT 4320

TGGGCATTTA AGGGCCAATT TACAGGAATT AGACGATTTT CTTAGTATTT TTTCATGTGT 4380

TAGCCCATAA ATATTTTGTT GATTTGACTT TTAGAGTCTA AACTTCTCAT GTATATTAAG 4440

AGATATTTAT GCTTGGTTAA TTGAATCGAA CTAGGAATAG AGAAATTCCT ACTTGGATCT 4500

TAATATTTCT CTCTCTTTGA TTTGGAAAAT TCTAGGAAGT TGCTTTCAAT GGAATTAAAA 4560

TCATCAATCT CTTGTATGTA AGAAACATAC TTATATTCAT GAATAGATAT GTTTAGGGTC 4620

TAATAATGAA TTATCACAAT TTTTTCTACT TTTTCTTGTC AGAGTCCTGC CTTTTTCTTT 4680

TTCTTTTTTA ACTTTGGTCT CTGCTTTTGT CTACATGATG ATAAGGTTGG TGGACCTAGC 4740

TGGAAATGTG ATGGAAATAG CTAGTAAAAG AAAGAACTTT GCATTTTCTG TTTTCTTAAA 4800

AACTGATAAA TTACATAACT TGTGGCAATT TGTCCATTTT CATACTGAGA GATATTTCTA 4860

TTTTTTTTGG ATAT**ATG**GCT TATGCTGCTG TTACTTCCCT TATGAGAACC ATACATCAAT 4920

CAATGGAACT TACTGGATGT GATTTGCAAC CGTTTTATGA AAAGCTCAAA TCTTTGAGAG 4980

CTATTCTGGA GAAATCCTGC AATATAATGG GCGATCATGA GGGGTTAACA ATCTTGGAAG 5040

TTGAAATCAT AGAGGTAGCA TACACAACAG AAGATATGGT TGACTCGGAA TCAAGAAATG 5100

TTTTTTTAGC ACGGAATGTG GGGAAAAGAA GCAGGGCTAT GTGGGGGATT TTTTTCGTCT 5160

TGGAACAAGC ACTAGAATGC ATTGATTCCA CCGTGAAACA GTGGATGGCA ACATCGGACA 5220

GCATGAAAGA TCTAAAACCA CAAACTAGCT CACTTGTCAG TTTACCTGAA CATGATGTTG 5280

AGCAGCCCGA GAATATAATG GTTGGCCGTG AAAATGAATT TGAGATGATG CTGGATCAAC 5340

TTGCTAGAGG AGGAAGGGAA CTAGAAGTTG TCTCAATCGT AGGGATGGGA GGCATCGGGA 5400

AAACAACTTT GGCTGCAAAA CTCTATAGTG ATCCTTACAT TATGTCTCGA TTTGATATTC 5460

GTGCAAAAGC AACTGTTTCA CAAGAGTATT GTGTGAGAAA TGTACTCCTA GGCCTTCTTT 5520

CTTTGACAAG TGATGAACCT GATTATCAGC TAGCGGACCA ACTGCAAAAG CATCTGAAAG 5580

GCAGGAGATA CTTGGTAGTC ATTGATGACA TATGGACTAC AGAAGCTTGG GATGATATAA 5640

AACTATGTTT CCCAGACTGC GATAATGGAA GCAGAATACT CCTGACTACT CGGAATGTGG 5700

AAGTGGCTGA ATATGCTAGC TCAGGTAAGC CTCCTCATCA CATGCGCCTC ATGAATTTTG 5760

ACGAAAGTTG GAATTTACTA CACAAAAAGA TCTTTGAAAA AGAAGGTTCT TATTCTCCTG 5820

AATTTGAAAA TATTGGGAAA CAAATTGCAT TAAAATGTGG AGGGTTACCT CTAGCAATTA 5880

CTTTGATTGC TGGACTTCTC TCCAAAATCA GTAAAACATT GGATGAGTGG CAAAATGTTG 5940

CGGAGAATGT ACGTTCGGTG GTAAGCACAG ATCTTGAAGC AAAATGCATG AGAGTGTTGG 6000

CTTTGAGTTA CCATCACTTG CCTTCTCACC TAAAACCGTG TTTTCTGTAT TTTGCAATTT 6060

TCGCAGAGGA TGAACGGATT TATGTAAATA AACTTGTTGA GTTATGGGCC GTAGAGGGGT 6120

TTTTGAATGA AGAAGAGGGA AAAAGCATAG AAGAGGTGGC AGAAACATGT ATAAACGAAC 6180

TTGTAGATAG AAGTCTAATT TCTATCCACA ATGTGAGTTT TGATGGGGAA ACACAGAGAT 6240

GTGGAATGCA TGATGTGACC CGTGAACTCT GTTTGAGGGA AGCTCGAAAC ATGAATTTTG 6300

TGAATGTTAT CAGAGGAAAG AGTGATCAAA ATTCATGTGC ACAATCCATG CAGTGTTCCT 6360

TTAAGAGTCG AAGTCGGATC AGTATCCATA ATGAGGAAGA ATTGGTTTGG TGTCGTAACA 6420

GCGAGGCTCA TTCTATCATC ACGTTGTGTA TATTCAAATG CGTCACACTG GAATTGTCTT 6480

TCAAGCTAGT AAGAGTACTA GATCTTGGTT TGACTACATG CCCAATTTTT CCCAGTGGAG 6540

TACTTTCTCT AATTCATTTG AGATACCTAT CTTTGCGTTT TAATCCTCGC TTACAGCAGT 6600

ATCGAGGATC GAAAGAAGCT GTTCCCTCAT CAATAATAGA CATTCCTCTA TCGATATCAA 6660

GCCTATGCTA TCTGCAAACT TTTAAACTTT ACCATCCATT TCCCAATTGT TATCCTTTCA 6720

TATTACCATC GGAAATTTTG ACAATGCCAC AATTGAGGAA GCTGTGTATG GGCTGGAATT 6780

ACTTGCGGAG TCATGAGCCT ACAGAGAACA GATTGGTTTT GAAAAGTTTG CAATGCCTCA 6840

ATGAATTGAA TCCTCGGTAT TGTACAGGGT CTTTTTTAAG ACTATTTCCC AATTTAAAGA 6900

AGTTGGAAGT ATTTGGCGTC AAAGAGGACT TTCGCAATCA CAAGGACCTG TATGATTTTC 6960

GCTACTTATA TCAGCTCGAG AAATTGGCAT TTAGTACTTA TTATTCATCT TCTGCTTGCT 7020

TTCTAAAAAA CACTGCACCT TTAGGTTCTA CTCCGCAAGA TCCTCTGAGG TTTCAGATGG 7080

AAACATTGCA CTTAGAGACT CATTCCAGGG CAACTGCACC TCCAACTGAT GTTCCAACTT 7140

TCCTCTTACC TCCTCCGGAT TGTTTTCCAC AAAACCTTAA GAGTTTAACT TTTAGCGGAG 7200

ATTTCTTTTT GGCATGGAAG GATTTGAGCA TTGTTGGTAA ATTACCCAAA CTCGAGGTCC 7260

TTCAACTATC ACACAATGCC TTCAAAGGCG AGGAGTGGGA AGTAGTTGAG GAAGGGTTTC 7320

CTCACTTGAA GTTCTTGTTT CTGGATAGCA TATACATTCG GTACTGGAGA GCTAGTAGTG 7380

ATCACTTTCC ATACCTTGAA CGACTTTTTC TTAGCGATTG CTTTTATTTG GATTCAATCC 7440

CTCGAGATTT TGCAGATATA ACCACACTAG CTCTTATTGA TATATTTCGC TGCCAACAAT 7500

CTGTTGGGAA TTCCGCCAAG CAAATTCAAC AGGACATTCA AGACAACTAT GGAAGCTCTA 7560

TCGAGGTCCA TACTCGTTAT CTTTA***GTAAG ACATCTTCTT CCTTGATTTA CAACAATATT*** 7620

TAACTCATCA TCATAGTAAA CTCGATAATA ATCTGGATAA TAGCTTTAGT AAGTCAAATT 7680

GCACCAATTC AACAAAAGTT CTTGATGCTG TCATTGTGAT TGATTCGAAT CCTTCCAATA 7740

TTGTGTAACT TGTTATACTT GCATGTTCAT TCTTGATTTT GGGAAGTGTA ACATTTCCAT 7800

TTTTCATCTT GATTTTGGGA AGTCGAAATG GAGCATTTTT GGTAGTG**TGA** CAACAGATGA 7860

AGATGATGAT GATAGTGTGA CAACAGATGA AGATGAAGAT GAAGACTTTG AGAAAGAAGT 7920

TGCTTCTTGC GGCAATAATG T***GTAAGTTCT TATACCTGCA TGCTCATTCT TGCTATAATG*** 7980

TTCTCTTGTT CCTTAATTAT GGGACATCTA ACATATTATT TTCCATTTTT TGCATCTTTT 8040

TTTTTTCCTG CAGCGTGTAG TTAAGGTGTT CTGAGGACTA GCCAGTTCTC TGAAATAAAT 8100

GTCAAATCAG AAGCCAAATG TGTGAGTGTT TGTTTTGTTC GTTTTCATTT TTTCTGCATA 8160

AGGTGGCAGG ATGATTGCAA ATGGCTTGTA ATTTAATTGT ATATGATATT TCGTATAGCC 8220

ATTTGCCAGT GGTTTTTTAG ATACTCCAAA TTTTATGTAC ATACATAATG GTATAGGCCA 8280

GAACAGGCTC CATATATAAC GTGTGTTTCC TTTCTTGGGA GTCCTCAATC TACCTCGCAA 8340

AGGAAGACAG ACGGCTAAAT CAAGAAAGAA ATTTTTTTGA AAATCATGTG GCTAGTTGTT 8400

CAACTTTATA CAAGTTTATG TGCATACTTG TGCATACCCA AAGTTGAATA ACATAAACAT 8460

AAAATGAAGT CAAGTTAAAT GGCACATTTA TGTATTATGC CTTTTGAATT TCATTAATAG 8520

TGAAAATCCT GAATCATATT CAGATTCCAT CACTAATCGT TGAACCATGT TAATTTACTA 8580

TGTATTATCT AATGGATTTT TTTGCTATCT TATTTATAAT TGTTCAAAGT TTTGTTAATT 8640

ATCTTTAGCA TAATATCTGA TTATATTATT TTGATATACT TTCTCTATCC CTAATTACTT 8700

GTCCATTTTT GAATTGGCAC ACCTATTAAG AAAATAATTA TTGAAATAGT GAGTTTACCA 8760

TTTTACCCAT ATTAATTATG AAGTGGATGA ATTAAAAACT CAAGATTTTC AAAAAGTTCT 8820

ATTTTTTTCA AAGTAATAAA CTGACGGTAT AATAGGTAAA AAAAATTATT CTTTCTTGAT 8880

TTGTCAAAAT AAACAAATAA TTAGGAATAA TTAAAAAAAT GGATAAATAA TTAAAAACGG 8940

AGGGAGCAAT ATGTTATCTT TAGCCTAATA ATATCTGATT AATGGCCACC CTAATTGATT 9000

GGATAGGAGA GGATAGACTT GCTTCCAAGT AACCCAAAAT ATAAAAAGTT GACAAAAGGG 9060

TGCTAAATTC GAGACACATG TAGTACTTAT ATAATTCATG TGCGGACTCG TTCTTTTGTA 9120

GTACTCCCTC CGTTCTATTT TATACGTCAC ATTTTTACTT TATACTTTTA TTAAGAAATG 9180

ATGTAGTTTT ATCTTTCTAT TCTTATTTAA TGTTTTCTTA AGTCAATTTT ATAATAAATA 9240

ATGAATATAT TTTCAAGATT AATTAACTAC TCTATCAAGG GTATAATAGG TAAAATATGA 9300

TAATTTATAC ATAAATTTTA TAAAATGACA AGTATTGTGG TCCAACTATT TATAGAAAGA 9360

AATGATATAT AAAATGGGAC GGAGGGCGTT ATAAAGTTGA CTTAAGAAAA CATTAAATAA 9420

GGGTAGAAGG GTAAAATTAC ATTATTTCTT AATGTAAATG TAAAGTAAAA AGGTAACATA 9480

TAAAATGGAA AGGAGGGAGT AGTATTTTCT TGTTTTATTT TACGTGGCAC TCTATTCTCA 9540

TAATCCGTCT TTAAAAATGT CATTTTATTG TAATTGAAAA TAATTTAACT TAAAATTCTC 9600

CATCTACCCT TAATTAATGA AATGATTTAC AATTATATAA ATATATAAAA ATTGTTTTAG 9660

CCTATAATTT TCTAAAATCT TTTTTTTTCT CTTATACATC GTATTAAGTC AAACATAAAT 9720

GGAATGGACG GAGTATTTCT TTTATTTTTT TGTCACACCG CCCATATGTT TTCTCCCATC 9780

CCCCAGACCC CCACTATGTA TATTCACTCC TTAGTTGGAT CTGAATTTAG AGTTTAGAAG 9840

CTTCTATAAT AATTTTAGAT TAATATATAA TAATAATAAT AATAATTGAA CTTACAGTAT 9900

TAAATTTATG TGAATCTATA TATATTGTAT TGTAATTTTT TTAATTATAA TTTTAACCAA 9960

ATCAATAAAG CTATTCAGAT GTAAAAGTAT ATATTATGAT TTAACAACAA ATTTCTATAC 10020

GTCTTCCTAA GTTTTGATGC ATAATTTCCT AAAACTCATA AATTTCCAAG TGACTACTTC 10080

CAGTATTACA ATGAGAACTT ATGTTTCGTT ATGGATTTTC TTAGTGAATT AGTTTAATAA 10140

AATCAAAATG AAAAAAAATC ATGTTTTATA ACATAAAATT TTCATTGATT CATGCGAAAA 10200

AAAAACATCT AGTTCTTATA GTGTGAAAAC TATTGAACTT ATGGGATGTA GCTGTATGGA 10260

AGTTCATCAA GTGGTAGCTC CTTGTACGCA ACTAGTGCTA CTTTTTATTG ACTAAAAGTT 10320

ATTTTCTAG 10329

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA oligonucleotide RG3

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: *Gpa2* encoding sequence of S. tuberosum

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

GGAGGCATCG GGAAAACAAC 20

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA oligonucleotide RG4

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: *Gpa2* encoding sequence of S. tuberosum

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

TGCTAGAGGT AAYCCTCC 18

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA oligonucleotide RG5

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: *Gpa2* encoding sequence of S. tuberosum

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

GATATGGTTG ACTCGGAATC AAG 23

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA oligonucleotide RG6

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: *Gpa2* encoding sequence of S. tuberosum

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GAGTATGGAC CTCGATAGAG C 21

REFERENCES

Aarts, M.G.M., te Lintel Hekkert, B., Holub, E.B., Beynon, J.L., Stiekema, W.J. and Pereira, A. (1998). Identification of R-gene homologous DNA fragments genetically linked to disease loci in *Arabidopsis thaliana. Mol Plant-Microbe Interact.* **11**: 251-258.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**: 403-410.

Arntzen, F.K., Visser, J.H.M. and Hoogendoorn, J. (1994). Inheritance, level and origin of resistance to *Globodera pallida* in the potato cultivar 'Multa', derived from *S. tuberosum* spp. *andigena* CPC1673. *Fundam. Appll. Nemat.* **16**: 155-162.

Baker, B., Zambryski, P., Staskawicz, B., and Dinesh-Kumar, S. P. (1997). Signaling in plant-microbe interactions. *Science* **276**: 726-733.

Bakker J., Bouwman-Smits L., and Gommers, F. J. (1992). Genetic relationships between *Globodera pallida* pathotypes in Europe assessed by using two dimensional gel electrophoresis of proteins. *Fundam Appl Nematol* **15**: 481-490.

Bakker, J., Folkertsma, R.T., Rouppe van der Voort, J.N.A.M., de Boer, J.M. and Gommers, F. (1993). Changing concepts and molecular approaches in the management of virulence genes in potato cyst nematodes. *Annu. Rev. Phytopathol.* **31**: 169-190.

Ballvora, A., Hesselbach, J., Niewohner, J., leister, D., Salamini, F and gebhardt, C. (1995). Marker enrichment and high-resolution map of the segment of potato chromosome VII harbouring the nematode resistance gene *Gro1. Mol. Gen. Genet.* **249**: 82-90.

Barone, A., Ritter, E., Schachtschabel, U., Debener, T, Salamini, F. and Gebhardt, C. (1990). Localization by restriction length polymorphism mapping in potato of a major dominant gene conferring resistance to the potato cyst nematode *Globodera rostochiensis. Mol Gen. genet.* **224**: 177-182.

Bendahmane, A., Kanyuka, K., and Baulcombe, D. C. (1997). High-resolution genetical and physical mapping of the *Rx* gene for extreme resistance to potato virus X in tetraploid potato. *Theor. Appl. Genet.* **95**: 153-162.

Bent, A. F. Kunkel, B. N., Dahlbeck, D., Brown, K. L., Schmidt, R., Giraudat, J., Leung, J., and Staskawicz, B. J. (1994). *RPS2* of *Arabidopsis thaliana*: A leucine-rich repeat class of plant disease resistance genes. *Science* **265**: 1856-1860.

Burge, C. and Karlin, S. (1997). Prediction of complete gene structure in human genomic DNA. *J. Mol. Biol.* **268**: 78-94.

Büschges, R., Hollricher, K., Panstruga, R., Simons, G., Wolter, M., Frijters, A., van Daelen, R., van der Lee, T., Groenendijk, J., Topsch, S., Vos, P., Salamini, F. and Schultze-Lefert, P. (1997). The barley *Mlo* gene: a novel control element of plant pathogen resistance. *Cell* **88**: 695-705.

Chapman, S., Kavanagh, T. and Baulcombe, D. (1992). Potato virus X as a vector for gene expression in plants. *Plant J.* **2**: 549-557.

Chu, G. (1989). Pulsed field electrophoresis in contour-clamped homogeneous electric fields for the resolution of DNA by size or topology. *Electrophoresis* **10**: 290-295.

Crute, I. R., and Pink, D. A. C. (1996). Genetics and utilization of pathogen resistance in plants. *Plant Cell* **8**:1747-1755.

Dear, S. and Staden, R. (1991). A sequence assembly and editing program for efficient management of large projects. *Nucleic Acids Res.* **14**: 3907-3911.

De Jong, W., Forsyth, A., Leister, D., Gebhardt, C., and Baulcombe, D. C. (1997). A potato hypersensitive resistance gene against potato virus X maps to a resistance gene cluster on chromosome 5. *Theor. Appl. Genet.* **95**: 246-252.

Engebrecht, J., Brent, R. and Kaderbhai, M.A. (1997). Large-scale preparation of plasmid DNA. In: Current Protocols in Molecular Biology, Vol. 1 (Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K., eds). John Wiley & Sons, Inc. pp. 1.6.1-1.6.2.

Fenwick, D.W. (1940). Methods for recovery and counting of cysts of Heterodera schachtii from soil. *J Helminth* **18**:155-172

Folkertsma, R.T. (1997). genetic diversity of the potato cyst nematode in the Netherlands. PhD thesis. Agricultural Univ. Wageningen, The Netherlands.

Gebhardt, C., Mugniery, D., Ritter, E., Salamini, F. and Bonnel. (1993). Identification of RFLP markers closely linked to the *H1* gene conferring resistance to *Globodera rostochiensis* in potato. *Theor. Appl. Genet.* **85**: 541-544.

Goulden, M.G., Köhm, B.A., Santa Cruz, S., kavanagh, T.A. and baulcombe, D. (1993). A feature of the coat protein of potao virus X affects both induced virus resistance in potato and viral fitness. *Virology* **197**: 293-302.

Heilig, J.S., Lech, K and Brent, R. (1997). Large-scale preparation of plasmid DNA. In: Current Protocols in Molecular Biology, Vol. 1 (Ausubel, F.M., Brent, R., Kingston,

R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K., eds). John Wiley & Sons, Inc. pp. 1.7.1-1.7.3.

Heungens, K., Mugniery, D., van Montagu, M., Gheysen, G. and Niebel, A. (1996). A method to obtain disinfected *Globodera* infective juveniles directly from cysts. *Fundam. appl. Nemat.*, **19**, 91-93.

Johal, G.S. and Briggs, S.P. (1992). Reductase activity encoded by the Hm1 disease resistance gene in maize. *Science* **258**: 985-987.

Konieczny, A. and Ausubel, F.M. (1993). A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. *Plant J.* **4**: 403-410.

Kreike, C. M., De Koning, J. R. A., Vinke, J. H., Van Ooijen, J. W., and Stiekema, W. J. (1994). Quantitatively inherited resistance to *Globodera pallida* is dominated by one major locus in *Solanum spegazzinii*. *Theor. Appl. Genet.* **88**: 764-769.

Lawrence, G. J., Finnegan, E. J., Ayliffe, M. A., and Ellis, J. G. (1995). The *L6* gene for flax rust resistance is related to the Arabidopsis bacterial resistance gene *RPS2* and the tobacco viral resistance gene *N*. *Plant Cell* 7:1195-1206.

Lazo, G.R., Stein, P.A., and Ludwig, R.A. (1991). A DNA transformation-competent *Arabidopsis* genomic library in *Agrobacterium*. *Bio/Technology* **9**: 963-967.

Leister, D., Ballvora, A., Salamini F., and Gebhardt, C. (1996). A PCR-based approach for isolating pathogen resistance genes from potato with potential for wide application in plants. *Nature Genetics* **14**: 421-429.

Leonards-Schippers, C., Gieffers, W., Salamini, F., and Gebhardt, C. (1992). The *R1* gene conferring race-specific resistance to *Phytophthora infestans* in potato is located on potato chromosome V. *Mol. Gen. Genet.* **233**: 378-383.

Liu, Y-G and Whittier. (1994). Rapid prparation of megabase plant DNA from nuclei in agarose plugs and microbeads. *Nucl. Acids Res.* **22**: 2168-2169.

Manly, K. F. (1994). Establishing genetic linkage using recombinant inbred lines with an abnormal segregation ratio. *Genetics* **136**:1434.

Maruzyk, R. and Sergeant, A. (1980). A simple method for dialysis of small volume samples. *Anal. Biochem.* **105**: 403-404.

Mindrinos, M., Katagiri, F., Yu, G. L., and Ausubel, F. M. (1994). The A. thaliana disease resistance gene *RPS2* encodes a protein a nucleotide-binding site and leucine - rich repeats. *Cell* **78**, 1089-1099.

Ochman, H., Mehdora, M.M., Garza, D. and Hartl, D.L. (1990). Amplification of flanking sequences by inverse PCR. In *PCR Protocols* (Innis, M.A., Gelfand, D.H., Sninsky, J.J. and White, T.J., eds). San Diego: Academis Press, pp.219-227.

Pineda, O., Bonierbale, M.W. and Plaisted, R.L. (1993). Identification of RFLP markers linked to the *H1* gene conferring resistance to the potato cyst nematode *Globodera rostochiensis. Genome* **36**: 153-156.

Rouppe van der Voort, J., Wolters, P., Folkertsma, R., Hutten, R., van Zandvoort, P., Vinke, H., Kanyuka, K., Bendahmane, A., Jacobsen, E., Janssen, R., and Bakker, J. (1997a). Mapping of the cyst nematode resistance locus *Gpa2* in potato using a strategy based on comigrating AFLP markers. *Theor. Appl. Genet.* **95**: 874-880.

Rouppe van der Voort J.N.A.M., Van Zandvoort P., Eck H.J. van, Folkertsma, F.T., Hutten, R.C.B., Draaistra J., Gommers F.J., Jacobsen E., Helder J. and Bakker J. (1997b). Allele specificity of comigrating AFLP markers used to align genetic maps from different potato genotypes. *Mol Gen Genet* **255**: 438-447.

Rouppe van der Voort, J., Lindeman, W., Folkertsma, R., Hutten, R., Overmars, H., van der Vossen, E., Jacobsen, E., and Bakker, J. (1998). A QTL for broad-spectrum resistance to cyst nematode species (genus *Globodera*) maps to a resistance gene cluster in potato. *Theor. Appl. Genet.* **96**: 654-661.

Sambrook J., Fritsch E.F. and Maniatis T. (1989). Molecular cloning: A Laboratory Manual (second edition). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Säll, T. and Nilsson, N. O. (1994). The robustness of recombination frequency estimates in intercrosses with dominant markers. *Genetics* **137**:589-596.

Stam, P. (1993). Construction of intergrated genetic linkage maps by menas of a new computer package JoinMap. *P. Journal* **3:** 739-744.

Staskawicz, B. J., Ausubel, F. M., Baker, B. J., Ellis, J. G. and Jones, J. D. G. (1995). Molecular genetics of plant-disease resistance. *Science* **268**:661-667.

Suiter, K.A., Wendel, J. F., and Chase, J. S. (1983). LINKAGE-1: a Pascal computer program for the detection and analysis of genetic linkage. *J. Hered.* **74**:203-204.

Tanksley, S.D., Ganal, M.W., Prince, J.P., de Vicente, M.C., Bonierbale, M.W., Broun, P., Fulton, T.M., Giovannoni, J.J., Grandillo, S., Martin, G.B., Messeguer, R., Miller, J.C., Miller, L., Paterson, A.H., Pineda, O., Röder, M.S., Wing, R.A.,

Wu, W. and Young, N.D. (1992). High density molecular linkage maps of the tomato and potato genomes. *Genetics* **132**: 1141-1160.

Van Engelen, F.A., Molthoff, J.W., Conner, A.J, Nap, J-P., Pereira, A. and Stiekema, W.J. (1995). pBIBPLUS: an improved plant transformation vector based on pBIN19. *Transgenic Research* 4: 288-290.

Visser, R.G.F. (1991). Regeneration and transformation of potato by *Agrobacterium tumefaciens*. In: Plant tissue culture manual. Kluwer Academic Publishers, Dordrecht Boston London, B5:1-9.

Whitham, S. Dinesh-Kumar, S.P., Choi, D., Hehl, R., Corr, C. and Baker, B. (1994). The product of the tobacco mosaic virus resistance gene N - similarity to Toll and the interleukin-1 receptor. *Cell* **78**: 1101-1115.

Young, N., and Tanksley, S. D. (1992). Restriction fragment length polymorphism maps and the concept of graphical genotypes. *Theor. Appl. Genet.* 77:95-101.

CLAIMS

1. A recombinant nucleic acid sequence providing resistance to infection by a phytopathogenic nematode of the genus *Globodera* when introduced into a host plant, said host plant prior to introduction being susceptible to infection with the phytopathogenic nematode, said introduction occurring in such a way that said nucleic acid sequence is expressed in the host plant, the nucleic acid sequence being that of SEQ ID NO.1.

2. A recombinant nucleic acid sequence being a homologue of the nucleic acid sequence according to claim 1, said homologue also providing the resistance, said homologue being a nucleic acid sequence encoding the amino acid sequence of SEQ ID NO.1.

3. A recombinant nucleic acid sequence being a homologue of the nucleic acid sequence according to claim 1, said homologue also providing the resistance, said homologue being a nucleic acid sequence exhibiting more than 70% homology at nucleic acid level with SEQ ID NO. 1.

4. A recombinant nucleic acid sequence being a homologue of the nucleic acid sequence according to claim 1, said homologue also providing the resistance, said homologue being a nucleic acid sequence exhibiting more than 75% homology at nucleic acid level with SEQ ID NO. 1.

5. A recombinant nucleic acid sequence being a homologue of the nucleic acid sequence according to claim 1, said homologue also providing the resistance, said homologue being a nucleic acid sequence exhibiting more than 80% homology at nucleic acid level with SEQ ID NO. 1.

6. A recombinant nucleic acid sequence being a homologue of the nucleic acid sequence according to claim 1, said homologue also providing the resistance, said homologue being a nucleic acid sequence exhibiting more than 85% homology at nucleic acid level with SEQ ID NO. 1.

7. A recombinant nucleic acid sequence being a homologue of the nucleic acid sequence according to claim 1, said homologue also providing the resistance, said homologue being a nucleic acid sequence exhibiting more than 90% homology at nucleic acid level with SEQ ID NO. 1.

8. A recombinant nucleic acid sequence being a homologue of the nucleic acid sequence according to claim 1, said homologue also providing the resistance, said homologue being a nucleic acid sequence exhibiting more than 95% homology at nucleic acid level with SEQ ID NO. 1.

9. A recombinant nucleic acid sequence being a homologue of the nucleic acid sequence according to claim 1 or being a homologue according to any of the claims 2-8, said homologue also providing the resistance, said homologue being a nucleic acid sequence capable of hybridising under normal to stringent conditions to the nucleic acid sequence of SEQ ID NO. 1.

10. A recombinant nucleic acid sequence being a homologue of the nucleic acid sequence according to claim 1 or being a homologue according to any of claims 2-9, said homologue also providing the resistance, said homologue being a nucleic acid sequence encoding a deletion, insertion or substitution mutant of the amino acid sequence of SEQ ID NO. 1.

11. A recombinant nucleic acid sequence being a homologue of the nucleic acid sequence according to claim 1 or being a homologue according to any of claims 2-10, said homologue also providing the resistance, said homologue being a nucleic acid sequence encoding a deletion, insertion or substitution variant as occurs in nature of the amino acid sequence of SEQ ID NO. 1.

12. A recombinant nucleic acid sequence according to any of the preceeding claims, said nucleic acid sequence further comprising at least one intron.

13. A recombinant nucleic acid sequence according to claim 12 comprising at least one intron of SEQ ID NO. 2.

14. A recombinant nucleic acid according to any of the preceeding claims being the genomic insert of pBINRGH2.

15. A recombinant nucleic acid sequence according to any of the preceeding claims, said nucleic acid sequence being identical to that present in the genetic material of a species of the family Solanacae, preferably a species of the genus *Solanum*.

16. A recombinant nucleic acid sequence according to any of the preceeding claims, said nucleic acid sequence being identical to that present in the genetic material of a potato, preferably on chromosome 4, 5, 6, 7, 9, 11 or 12.

17. A recombinant nucleic acid sequence according to any of the preceeding claims, said nucleic acid sequence being identical to that present in the genetic material of potato locus *Gpa2*.

18. A recombinant nucleic acid sequence being a homologue of the nucleic acid sequence according to claim 1, said homologue also providing the resistance, said homologue being a fragment of the nucleic acid sequence according to any of claims 14-17.

19. A genetic construct comprising a nucleic acid sequence according to any of the preceeding claims said sequence being operably linked to a regulatory region for expression.

20. A genetic construct according to claim 19 wherein the regulatory region is a *Gpa2* regulatory region.

21. A genetic construct according to any of claims 19 or 20 wherein the regulatory region corresponds to that present in the sequence of nucleotides 1-4874 of SEQ ID NO. 3.

22. A genetic construct according to any of claims 19-21, wherein the regulatory region corresponds to that of nucleotides 1-4874 of SEQ ID NO.3.

23. A genetic construct according to any of the preceeding claims 19-22, wherein the regulatory region comprises a promoter functionally connected to a nucleic acid sequence as defined in any of the claims 1-18, said promoter being able to control the transcription of said nucleic acid sequence in a plant cell.

24. A vector which carries a nucleic acid according to any of the claims 1-18, or a genetic construct according to any of the claims 19-23.

25. A vector according to claim 24 capable of replicating in a host organism.

26. A vector capable of expressing the nucleic acid according to any of the claims 1-19, or a genetic construct according to any of the claims 19-23.

27. A vector according to any of claims 24-26 constructed to function in a host organism selected from the group consisting of a micro-organism, plant cell, plant, seed, seedling, plant part and protoplast.

28. A vector according to any of claims 24-27 constructed to function in a host organism selected from the group consisting of a micro-organism, plant cell, seed, seedling, plant part and protoplast.

29. A vector according to any of claims 24-28 constructed to function in a host organism selected from the group consisting of a micro-organism, plant cell, plant part and protoplast.

30. A vector according to any of claims 24-29 constructed to function in a host organism selected from the group consisting of a plant, plant cell, plant part, seed, seedling and protoplast.

31. A host organism selected from the group consisting of a micro-organism, plant cell, plant, seed, seedling, plant part and protoplast, harbouring a vector as defined in any of claims 24-30 and/or a genetic construct according to any of the claims 19-23.

32. A host organism selected from the group consisting of a micro-organism, plant cell, seed, seedling, plant part and protoplast, harbouring a vector as defined in any of claims 24-30 and/or a genetic construct according to any of the claims 19-23.

33. A host organism selected from the group consisting of a micro-organism, plant cell, plant part and protoplast, harbouring a vector as defined in any of claims 24-30 and/or a genetic construct according to any of the claims 19-23.

34. A host organism selected from the group consisting of a plant cell, plant, seed, seedling, plant part and protoplast, harbouring a vector as defined in any of claims 24-30 and/or a genetic construct according to any of the claims 19-23.

35. A host organism according to any preceeding claim 31-34 which is capable of replicating or expressing the nucleic acid sequence or the genetic construct of the vector and/or a genetic construct according to any of the claims 19-23.

36. A process for producing a genetically transformed or transfected host organism having increased resistance to phytopathogenic nematodes of the genus *Globodera* as compared to the host organism prior to the transformation, said process comprising transferring a genetic construct according to any of the claims 19-23 and/or a vector according to any of claims 24-30 into the host organism so that it's genetic material comprises the genetic construct and subsequently regenerating the host organism into a genetically transformed plant part.

37. A process according to claim 36 for producing a genetically transformed plant having increased resistance to phytopathogenic nematodes of the genus *Globodera* as compared to a corresponding plant prior to the transformation, said process comprising transferring a genetic construct according to any of the claims 19-23 and/or a vector according to any of claims 24-30 into the host organism so that it's genetic material comprises the genetic construct and/or a vector according to any of claims 19-23 and subsequently regenerating the host organism into a genetically transformed plant, said host organism being selected from the group consisting of a plant cell, plant, seed, seedling, plant part and protoplast of the plant type to be rendered resistant.

38. A process according to claim 36 or 37 wherein said nematodes are selected from the group consisting of *Globodera pallida* and *Globodera rostochiensis.*

39. A process according to any of claims 36-38, wherein said host organism to be transformed is selected from a plant type of the family Solanacae.

40. A process according to any of claims 36-39, wherein said host organism to be transformed is selected from a plant type of the genus *Solanum.*

41. A process according to any of claims 36-40, wherein said host organism to be transformed is selected from a plant type of the species *Solanum tuberosum.*

42. A process for isolating or detecting a nucleic acid sequence according to any of claims 1-18, comprising the screening of genomic nucleic acid of a plant with a nucleic acid sequence according to any of claims 1-18 or a fragment thereof as probe or primer, said probe or primer being at least 16 nucleotides in length, the identification of positive clones which hybridize to said probe and the isolation of said positive clones and the isolation of the nucleic acid sequence therefrom.

43. A process for isolating or detecting a nucleic acid sequence according to claims 1-18, comprising the screening of a genomic library of a plant with a nucleic sequence according to seq id no 3 or a fragment thereof as probe, said probe or primer being at least 16 nucleotides in length, the identification of positive clones which hybridize to said probe or primer and the isolation of said positive clones and the isolation of the nucleic acid sequence therefrom.

44. A process for isolating or detecting a nucleic acid sequence according to claims 1-18, comprising the screening of a cDNA library of a plant with the encoding portion of a nucleic acid sequence according to any of claims 1-18 or a fragment thereof as probe or primer, said probe or primer being at least 16 nucleotides in length, the identification of positive clones which hybridize to said probe or primer and the isolation of said positive clones and the isolation of the nucleic acid sequence therefrom.

45. A process for isolating or detecting a nucleic acid sequence according to claims 1-18, comprising the screening of a cDNA library of a plant with the encoding portion of a nucleic acid sequence according to SEQ ID NO. 1 or a fragment thereof as probe or primer, said probe or primer being at least 16 nucleotides in length, the identification of positive clones which hybridize to said probe and the isolation of said positive clones and the isolation of the nucleic acid sequence therefrom.

46. A process according to any of claims 42-45, wherein the probe is comprised within the sequence of SEQ ID NO.1, SEQ ID NO.2 or SEQ ID NO.3.

47. A process for isolating or detecting a nucleic acid sequence according to any of claims 1-18, using a nucleic acid amplification reaction such as the Polymerase Chain Reaction and at least one primer corresponding to or being complementary to the nucleic acid sequence according to any of claims 1-18 or a fragment thereof as primer, said primer being at least 16 nucleotides in length.

48. A process for isolating or detecting a nucleic acid sequence according to any of claims 1-18, using a nucleic acid amplification reaction such as the Polymerase Chain Reaction and at least one primer corresponding to or being complementary to the nucleic acid sequence of of SEQ ID NO.1, SEQ ID NO.2 or SEQ ID NO.3 or a fragment thereof as primer, said primer being at least 16 nucleotides in length.

49. A process according to any of claims 42-48 wherein said probe or primer comprises a nucleic acid sequence encoding at least part of the amino acid sequence of the NBS sequence of *Gpa2*.

50. A process according to any of claims 42-49, wherein said probe or primer comprises a nucleic acid sequence encoding at least part of the amino acid sequence of the NBS sequence of *Gpa2*, at least part having the following sequence GGIGKTT or GGLPLA.

51. A process according to any of claims 42-50, wherein said probe or primer comprises parts of the NBS sequence of *Gpa2* and at least one codon of a 5' and/or 3' overhanging

portion corresponding to the respective 5' and/or 3' adjacent amino acids of the specified NBS sequence of *Gpa2*.

52. A process according to any of claims 42-51, wherein said probe or primer comprises parts of the NBS sequence of *Gpa2* and at least one codon of a 5' and/or 3' overhanging portion corresponding to the respective 5' and/or 3' adjacent amino acids of the specified NBS sequence of *Gpa2* of SEQ ID NO.1.

53. A process according to any of claims 42-52, wherein said probe or primer corresponds to a sequence selected from SEQ ID NO.4, SEQ ID NO.5, SEQ ID NO.6 and/or SEQ ID NO.7.

54. A polypeptide having an amino acid sequence provided in SEQ ID NO.1 or being a homologue of said amino acid sequence, said homologue being a substitution, insertion or deletion mutant possessing nematode resistance against a nematode of the genus *Globodera*.

55. A polypeptide encoded by a sequence according to any of the claims 1-18.

56. A process for producing a polypeptide having an amino acid sequence provided in SEQ ID NO.1 or a homologue of said amino acid sequence, said homologue being a substitution, insertion or deletion mutant possessing nematode resistance against a nematode of the genus *Globodera,* said process comprising expressing a recombinant nucleic acid sequence according to any of the claims 1-18 or genetic construct according to any of claims 19-23 and optionally isolating said polypeptide, said expression occurring in a host organism according to any of claims 31-35.

57. A process for producing a polypeptide having an amino acid sequence provided in SEQ ID NO.1 or a homologue of said amino acid sequence, said homologue being a substitution, insertion or deletion mutant possessing nematode resistance against a nematode of the genus *Globodera*, said process comprising the expression of a recombinant nucleic acid sequence according to any of the claims 1-18 or genetic

construct according to any of claims 19-23 and isolating said polypeptide, said expression occurring in a host organism according to any of claims 31-35.

58. A nematicide composition comprising as active ingredient a polypeptide according to claim 54 or 55 or produced according to claim 56 or 57 or a host organism expressing such a polypeptide, such a host organism being defined in any of claims 31-35 in a formulation suitable for application as nematicide to a plant and optionally comprising other ingredients required for rendering the polypeptide suitable for application as a nematicide.

59. A nematicide composition according to claim 58 comprising the polypeptide in a slow release dosage form.

60. A nematicide composition according to 58 or 59 comprising instructions for application as nematicide.

61. A nucleic acid sequence comprising at least 16 contiguous nucleotides corresponding to or complementary to the *Gpa2* sequence, with the *proviso* that when such an oligonucleotide comprises part or all of the NBS encoding sequence, the oligonucleotide also comprises at least one codon of a 5' and/or 3' overhanging portion corresponding to the respective 5' and/or 3' adjacent amino acids of the specified NBS sequence of *Gpa2.*

62. A nucleic acid sequence according to claim 61, wherein the *Gpa2* sequence is comprised within the sequence of SEQ ID NO.1, 2 or 3.

63. A nucleic acid sequence according to claim 61 or 62, wherein sequence length is at least 50 nucleotides, suitably more than 100 nucleotides and is suitable for use as probe or primer in a nucleic acid assay.

64. A nucleic acid sequence according to any of claims 61-63, being selected from any of the sequences SEQ ID NOs. 4, 5, 6 and/or 7.

65. A combination of at least 2 primers according to any of claims 61-64.

66. Antibody raised against a polypeptide of claim 55 or a polypeptide produced by a process according to claim 56 or 57.

67. A diagnostic kit for assessing the presence of nematode resistance of a plant to infection by a phytopathogenic nematode of the genus *Globodera*, said kit comprising at least one nucleic acid sequence according to any of claims 61-64 as a probe or primer for screening of nucleic acid from a plant or plant part to be tested and/or a combination of primers according to claim 65 and/or an antibody according to claim 66.

68. A process for diagnosing whether a plant is resistant to a phytopathogenic nematode of the genus *Globodera*, said process comprising detecting the presence of a nucleic acid sequence according to any of claims 1-18, genetic construct according to any of claims 19-23, vector according to any of 24-30 or a polypeptide according to claim 55 in plant material of a plant to be tested.

69. A process for diagnosing whether a plant is resistant to a phytopathogenic nematode of the genus *Globodera*, said process comprising carrying out a process according to any of claims 42-53 and/or applying a diagnostic kit acording to claim 67.

70. A process for protecting plants said process comprising the introduction of the nucleic acid sequence according to any of claims 1-18, the genetic construct according to any of claims 19-23, the vector according to any of 24-30 in plant material of a plant to be protected.

Fig - 1

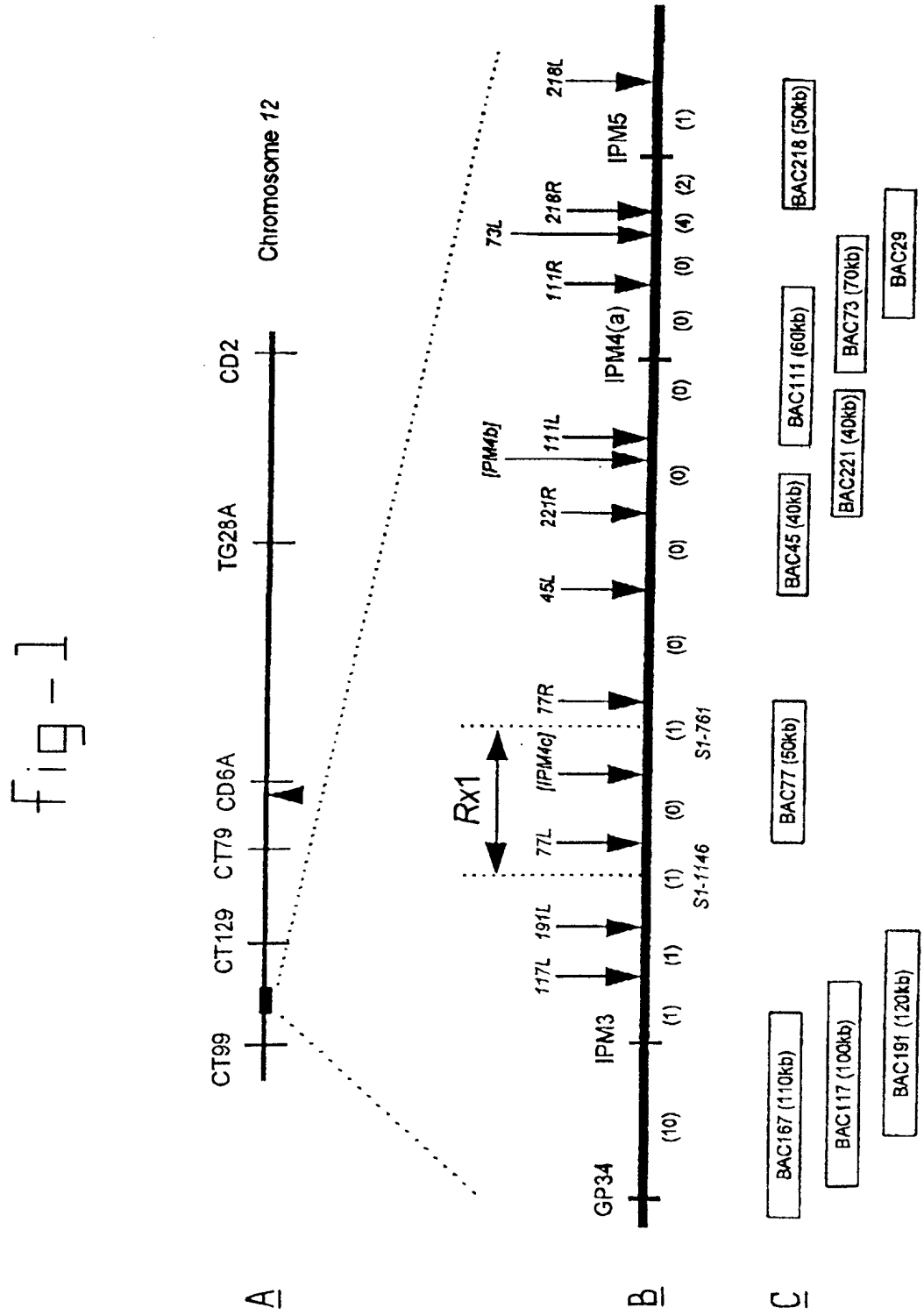
A
Chromosome 12
CD2
TG28A
CD6A
CT79
CT129
CT99
B
C
Rx1
GP34
IPM3
IPM4(a)
IPM5
117L
191L
77L
[IPM4c]
77R
45L
221R
[IPM4b]
111L
111R
73L
218R
218L
S1-1146
S1-761
(10)
(1)
(0)
(2)
(4)
BAC167 (110kb)
BAC117 (100kb)
BAC191 (120kb)
BAC77 (50kb)
BAC45 (40kb)
BAC111 (60kb)
BAC221 (40kb)
BAC73 (70kb)
BAC218 (50kb)
BAC29

fig-2

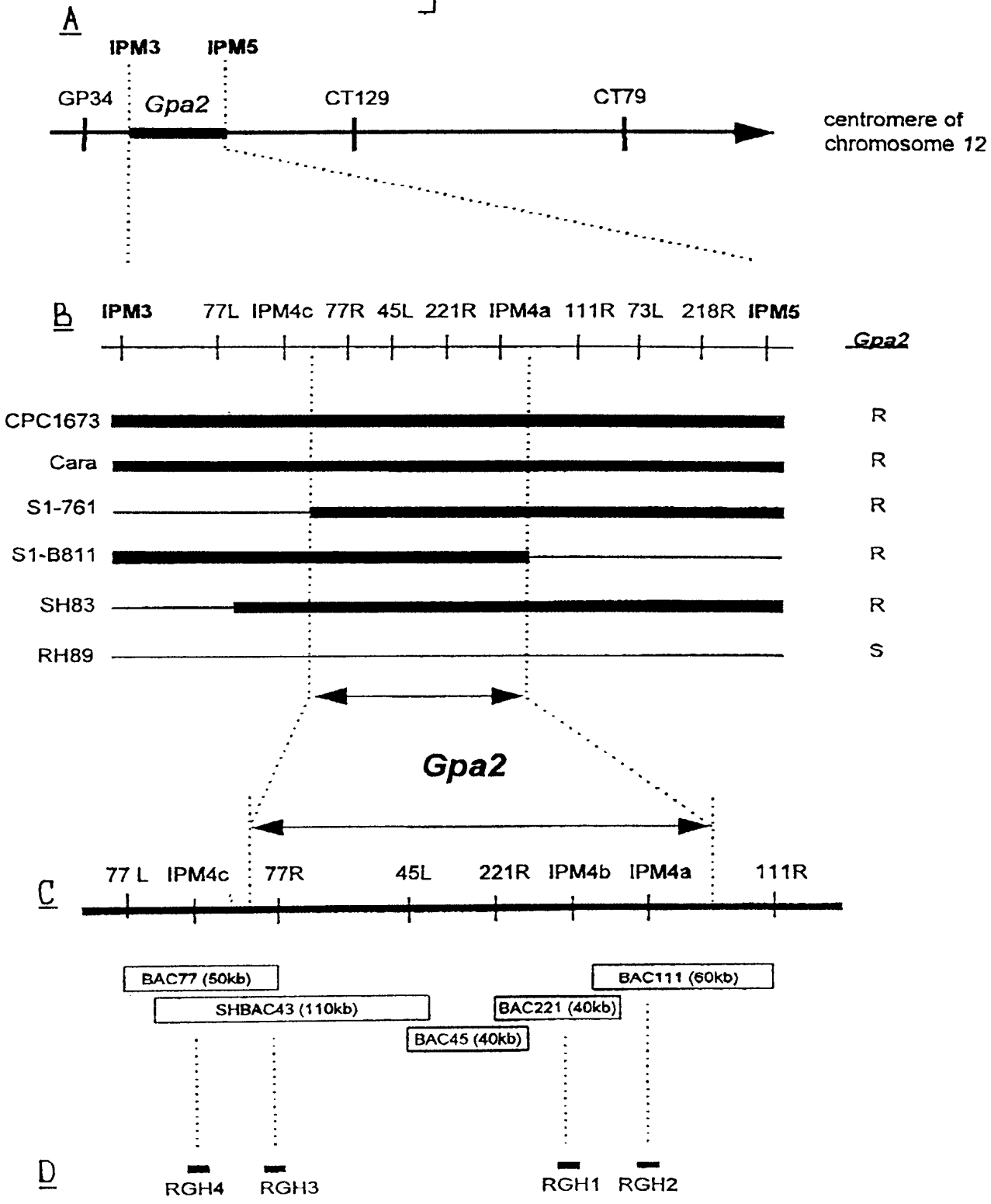
A
IPM3
IPM5
GP34
Gpa2
CT129
CT79
centromere of chromosome 12
B
IPM3
77L
IPM4c
77R
45L
221R
IPM4a
111R
73L
218R
IPM5
Gpa2
CPC1673
Cara
S1-761
S1-B811
SH83
RH89
R
R
R
R
R
S
Gpa2
C
77 L
IPM4c
77R
45L
221R
IPM4b
IPM4a
111R
BAC77 (50kb)
SHBAC43 (110kb)
BAC45 (40kb)
BAC221 (40kb)
BAC111 (60kb)
D
RGH4
RGH3
RGH1
RGH2

Fig. 3a (1)

```
atg gct tat gct gct gtt act tcc ctt atg aga acc ata cat caa tca   48
Met Ala Tyr Ala Ala Val Thr Ser Leu Met Arg Thr Ile His Gln Ser
  1               5                  10                  15

atg gaa ctt act gga tgt gat ttg caa ccg ttt tat gaa aag ctc aaa   96
Met Glu Leu Thr Gly Cys Asp Leu Gln Pro Phe Tyr Glu Lys Leu Lys
             20                  25                  30

tct ttg aga gct att ctg gag aaa tcc tgc aat ata atg ggc gat cat  144
Ser Leu Arg Ala Ile Leu Glu Lys Ser Cys Asn Ile Met Gly Asp His
         35                  40                  45

gag ggg tta aca atc ttg gaa gtt gaa atc ata gag gta gca tac aca  192
Glu Gly Leu Thr Ile Leu Glu Val Glu Ile Ile Glu Val Ala Tyr Thr
     50                  55                  60

aca gaa gat atg gtt gac tcg gaa tca aga aat gtt ttt tta gca cgg  240
Thr Glu Asp Met Val Asp Ser Glu Ser Arg Asn Val Phe Leu Ala Arg
 65                  70                  75                  80

aat gtg ggg aaa aga agc agg gct atg tgg ggg att ttt ttc gtc ttg  288
Asn Val Gly Lys Arg Ser Arg Ala Met Trp Gly Ile Phe Phe Val Leu
                 85                  90                  95

gaa caa gca cta gaa tgc att gat tcc acc gtg aaa cag tgg atg gca  336
Glu Gln Ala Leu Glu Cys Ile Asp Ser Thr Val Lys Gln Trp Met Ala
            100                 105                 110

aca tcg gac agc atg aaa gat cta aaa cca caa act agc tca ctt gtc  384
Thr Ser Asp Ser Met Lys Asp Leu Lys Pro Gln Thr Ser Ser Leu Val
        115                 120                 125

agt tta cct gaa cat gat gtt gag cag ccc gag aat ata atg gtt ggc  432
Ser Leu Pro Glu His Asp Val Glu Gln Pro Glu Asn Ile Met Val Gly
    130                 135                 140

cgt gaa aat gaa ttt gag atg atg ctg gat caa ctt gct aga gga gga  480
Arg Glu Asn Glu Phe Glu Met Met Leu Asp Gln Leu Ala Arg Gly Gly
145                 150                 155                 160

agg gaa cta gaa gtt gtc tca atc gta ggg atg gga ggc atc ggg aaa  528
Arg Glu Leu Glu Val Val Ser Ile Val Gly Met Gly Gly Ile Gly Lys
                165                 170                 175

aca act ttg gct gca aaa ctc tat agt gat cct tac att atg tct cga  576
Thr Thr Leu Ala Ala Lys Leu Tyr Ser Asp Pro Tyr Ile Met Ser Arg
            180                 185                 190

ttt gat att cgt gca aaa gca act gtt tca caa gag tat tgt gtg aga  624
Phe Asp Ile Arg Ala Lys Ala Thr Val Ser Gln Glu Tyr Cys Val Arg
        195                 200                 205
```

Fig. 3a (2)

```
aat gta ctc cta ggc ctt ctt tct ttg aca agt gat gaa cct gat tat   672
Asn Val Leu Leu Gly Leu Leu Ser Leu Thr Ser Asp Glu Pro Asp Tyr
    210                 215                 220

cag cta gcg gac caa ctg caa aag cat ctg aaa ggc agg aga tac ttg   720
Gln Leu Ala Asp Gln Leu Gln Lys His Leu Lys Gly Arg Arg Tyr Leu
225                 230                 235                 240

gta gtc att gat gac ata tgg act aca gaa gct tgg gat gat ata aaa   768
Val Val Ile Asp Asp Ile Trp Thr Thr Glu Ala Trp Asp Asp Ile Lys
                245                 250                 255

cta tgt ttc cca gac tgc gat aat gga agc aga ata ctc ctg act act   816
Leu Cys Phe Pro Asp Cys Asp Asn Gly Ser Arg Ile Leu Leu Thr Thr
            260                 265                 270

cgg aat gtg gaa gtg gct gaa tat gct agc tca ggt aag cct cct cat   864
Arg Asn Val Glu Val Ala Glu Tyr Ala Ser Ser Gly Lys Pro Pro His
        275                 280                 285

cac atg cgc ctc atg aat ttt gac gaa agt tgg aat tta cta cac aaa   912
His Met Arg Leu Met Asn Phe Asp Glu Ser Trp Asn Leu Leu His Lys
    290                 295                 300

aag atc ttt gaa aaa gaa ggt tct tat tct cct gaa ttt gaa aat att   960
Lys Ile Phe Glu Lys Glu Gly Ser Tyr Ser Pro Glu Phe Glu Asn Ile
305                 310                 315                 320

ggg aaa caa att gca tta aaa tgt gga ggg tta cct cta gca att act  1008
Gly Lys Gln Ile Ala Leu Lys Cys Gly Gly Leu Pro Leu Ala Ile Thr
                325                 330                 335

ttg att gct gga ctt ctc tcc aaa atc agt aaa aca ttg gat gag tgg  1056
Leu Ile Ala Gly Leu Leu Ser Lys Ile Ser Lys Thr Leu Asp Glu Trp
            340                 345                 350

caa aat gtt gcg gag aat gta cgt tcg gtg gta agc aca gat ctt gaa  1104
Gln Asn Val Ala Glu Asn Val Arg Ser Val Val Ser Thr Asp Leu Glu
        355                 360                 365

gca aaa tgc atg aga gtg ttg gct ttg agt tac cat cac ttg cct tct  1152
Ala Lys Cys Met Arg Val Leu Ala Leu Ser Tyr His His Leu Pro Ser
    370                 375                 380

cac cta aaa ccg tgt ttt ctg tat ttt gca att ttc gca gag gat gaa  1200
His Leu Lys Pro Cys Phe Leu Tyr Phe Ala Ile Phe Ala Glu Asp Glu
385                 390                 395                 400

cgg att tat gta aat aaa ctt gtt gag tta tgg gcc gta gag ggg ttt  1248
Arg Ile Tyr Val Asn Lys Leu Val Glu Leu Trp Ala Val Glu Gly Phe
                405                 410                 415
```

Fig. 3a (3)

```
ttg aat gaa gaa gag gga aaa agc ata gaa gag gtg gca gaa aca tgt   1296
Leu Asn Glu Glu Glu Gly Lys Ser Ile Glu Glu Val Ala Glu Thr Cys
            420                 425                 430

ata aac gaa ctt gta gat aga agt cta att tct atc cac aat gtg agt   1344
Ile Asn Glu Leu Val Asp Arg Ser Leu Ile Ser Ile His Asn Val Ser
        435                 440                 445

ttt gat ggg gaa aca cag aga tgt gga atg cat gat gtg acc cgt gaa   1392
Phe Asp Gly Glu Thr Gln Arg Cys Gly Met His Asp Val Thr Arg Glu
    450                 455                 460

ctc tgt ttg agg gaa gct cga aac atg aat ttt gtg aat gtt atc aga   1440
Leu Cys Leu Arg Glu Ala Arg Asn Met Asn Phe Val Asn Val Ile Arg
465                 470                 475                 480

gga aag agt gat caa aat tca tgt gca caa tcc atg cag tgt tcc ttt   1488
Gly Lys Ser Asp Gln Asn Ser Cys Ala Gln Ser Met Gln Cys Ser Phe
                485                 490                 495

aag agt cga agt cgg atc agt atc cat aat gag gaa gaa ttg gtt tgg   1536
Lys Ser Arg Ser Arg Ile Ser Ile His Asn Glu Glu Glu Leu Val Trp
            500                 505                 510

tgt cgt aac agc gag gct cat tct atc atc acg ttg tgt ata ttc aaa   1584
Cys Arg Asn Ser Glu Ala His Ser Ile Ile Thr Leu Cys Ile Phe Lys
        515                 520                 525

tgc gtc aca ctg gaa ttg tct ttc aag cta gta aga gta cta gat ctt   1632
Cys Val Thr Leu Glu Leu Ser Phe Lys Leu Val Arg Val Leu Asp Leu
    530                 535                 540

ggt ttg act aca tgc cca att ttt ccc agt gga gta ctt tct cta att   1680
Gly Leu Thr Thr Cys Pro Ile Phe Pro Ser Gly Val Leu Ser Leu Ile
545                 550                 555                 560

cat ttg aga tac cta tct ttg cgt ttt aat cct cgc tta cag cag tat   1728
His Leu Arg Tyr Leu Ser Leu Arg Phe Asn Pro Arg Leu Gln Gln Tyr
                565                 570                 575

cga gga tcg aaa gaa gct gtt ccc tca tca ata ata gac att cct cta   1776
Arg Gly Ser Lys Glu Ala Val Pro Ser Ser Ile Ile Asp Ile Pro Leu
            580                 585                 590

tcg ata tca agc cta tgc tat ctg caa act ttt aaa ctt tac cat cca   1824
Ser Ile Ser Ser Leu Cys Tyr Leu Gln Thr Phe Lys Leu Tyr His Pro
        595                 600                 605

ttt ccc aat tgt tat cct ttc ata tta cca tcg gaa att ttg aca atg   1872
Phe Pro Asn Cys Tyr Pro Phe Ile Leu Pro Ser Glu Ile Leu Thr Met
    610                 615                 620
```

Fig. 3a (4)

```
cca caa ttg agg aag ctg tgt atg ggc tgg aat tac ttg cgg agt cat   1920
Pro Gln Leu Arg Lys Leu Cys Met Gly Trp Asn Tyr Leu Arg Ser His
625                 630                 635                 640

gag cct aca gag aac aga ttg gtt ttg aaa agt ttg caa tgc ctc aat   1968
Glu Pro Thr Glu Asn Arg Leu Val Leu Lys Ser Leu Gln Cys Leu Asn
                645                 650                 655

gaa ttg aat cct cgg tat tgt aca ggg tct ttt tta aga cta ttt ccc   2016
Glu Leu Asn Pro Arg Tyr Cys Thr Gly Ser Phe Leu Arg Leu Phe Pro
            660                 665                 670

aat tta aag aag ttg gaa gta ttt ggc gtc aaa gag gac ttt cgc aat   2064
Asn Leu Lys Lys Leu Glu Val Phe Gly Val Lys Glu Asp Phe Arg Asn
        675                 680                 685

cac aag gac ctg tat gat ttt cgc tac tta tat cag ctc gag aaa ttg   2112
His Lys Asp Leu Tyr Asp Phe Arg Tyr Leu Tyr Gln Leu Glu Lys Leu
    690                 695                 700

gca ttt agt act tat tat tca tct tct gct tgc ttt cta aaa aac act   2160
Ala Phe Ser Thr Tyr Tyr Ser Ser Ser Ala Cys Phe Leu Lys Asn Thr
705                 710                 715                 720

gca cct tta ggt tct act ccg caa gat cct ctg agg ttt cag atg gaa   2208
Ala Pro Leu Gly Ser Thr Pro Gln Asp Pro Leu Arg Phe Gln Met Glu
                725                 730                 735

aca ttg cac tta gag act cat tcc agg gca act gca cct cca act gat   2256
Thr Leu His Leu Glu Thr His Ser Arg Ala Thr Ala Pro Pro Thr Asp
            740                 745                 750

gtt cca act ttc ctc tta cct cct ccg gat tgt ttt cca caa aac ctt   2304
Val Pro Thr Phe Leu Leu Pro Pro Pro Asp Cys Phe Pro Gln Asn Leu
        755                 760                 765

aag agt tta act ttt agc gga gat ttc ttt ttg gca tgg aag gat ttg   2352
Lys Ser Leu Thr Phe Ser Gly Asp Phe Phe Leu Ala Trp Lys Asp Leu
    770                 775                 780

agc att gtt ggt aaa tta ccc aaa ctc gag gtc ctt caa cta tca cac   2400
Ser Ile Val Gly Lys Leu Pro Lys Leu Glu Val Leu Gln Leu Ser His
785                 790                 795                 800

aat gcc ttc aaa ggc gag gag tgg gaa gta gtt gag gaa ggg ttt cct   2448
Asn Ala Phe Lys Gly Glu Glu Trp Glu Val Val Glu Glu Gly Phe Pro
                805                 810                 815

cac ttg aag ttc ttg ttt ctg gat agc ata tac att cgg tac tgg aga   2496
His Leu Lys Phe Leu Phe Leu Asp Ser Ile Tyr Ile Arg Tyr Trp Arg
            820                 825                 830
```

Fig. 3a (5)

```
gct agt agt gat cac ttt cca tac ctt gaa cga ctt ttt ctt agc gat   2544
Ala Ser Ser Asp His Phe Pro Tyr Leu Glu Arg Leu Phe Leu Ser Asp
        835                 840                 845

tgc ttt tat ttg gat tca atc cct cga gat ttt gca gat ata acc aca   2592
Cys Phe Tyr Leu Asp Ser Ile Pro Arg Asp Phe Ala Asp Ile Thr Thr
    850                 855                 860

cta gct ctt att gat ata ttt cgc tgc caa caa tct gtt ggg aat tcc   2640
Leu Ala Leu Ile Asp Ile Phe Arg Cys Gln Gln Ser Val Gly Asn Ser
865                 870                 875                 880

gcc aag caa att caa cag gac att caa gac aac tat gga agc tct atc   2688
Ala Lys Gln Ile Gln Gln Asp Ile Gln Asp Asn Tyr Gly Ser Ser Ile
                885                 890                 895

gag gtc cat act cgt tat ctt tat cga aat gga gca ttt ttg gta gtg   2736
Glu Val His Thr Arg Tyr Leu Tyr Arg Asn Gly Ala Phe Leu Val Val
            900                 905                 910

tga                                                              2739
 *
```

Fig. 3b (1)

ATGGCTTATG CTGCTGTTAC TTCCCTTATG AGAACCATAC ATCAATCAAT GGAACTTACT 60

GGATGTGATT TGCAACCGTT TTATGAAAAG CTCAAATCTT TGAGAGCTAT TCTGGAGAAA 120

TCCTGCAATA TAATGGGCGA TCATGAGGGG TTAACAATCT TGGAAGTTGA AATCATAGAG 180

GTAGCATACA CAACAGAAGA TATGGTTGAC TCGGAATCAA GAAATGTTTT TTTAGCACGG 240

AATGTGGGGA AAAGAAGCAG GGCTATGTGG GGGATTTTTT TCGTCTTGGA ACAAGCACTA 300

GAATGCATTG ATTCCACCGT GAAACAGTGG ATGGCAACAT CGGACAGCAT GAAAGATCTA 360

AAACCACAAA CTAGCTCACT TGTCAGTTTA CCTGAACATG ATGTTGAGCA GCCCGAGAAT 420

ATAATGGTTG GCCGTGAAAA TGAATTTGAG ATGATGCTGG ATCAACTTGC TAGAGGAGGA 480

AGGGAACTAG AAGTTGTCTC AATCGTAGGG ATGGGAGGCA TCGGGAAAAC AACTTTGGCT 540

GCAAAACTCT ATAGTGATCC TTACATTATG TCTCGATTTG ATATTCGTGC AAAAGCAACT 600

GTTTCACAAG AGTATTGTGT GAGAAATGTA CTCCTAGGCC TTCTTTCTTT GACAAGTGAT 660

GAACCTGATT ATCAGCTAGC GGACCAACTG CAAAAGCATC TGAAAGGCAG GAGATACTTG 720

GTAGTCATTG ATGACATATG GACTACAGAA GCTTGGGATG ATATAAAACT ATGTTTCCCA 780

GACTGCGATA ATGGAAGCAG AATACTCCTG ACTACTCGGA ATGTGGAAGT GGCTGAATAT 840

GCTAGCTCAG GTAAGCCTCC TCATCACATG CGCCTCATGA ATTTTGACGA AAGTTGGAAT 900

TTACTACACA AAAAGATCTT TGAAAAAGAA GGTTCTTATT CTCCTGAATT TGAAAATATT 960

GGGAAACAAA TTGCATTAAA ATGTGGAGGG TTACCTCTAG CAATTACTTT GATTGCTGGA 1020

CTTCTCTCCA AAATCAGTAA AACATTGGAT GAGTGGCAAA ATGTTGCGGA GAATGTACGT 1080

TCGGTGGTAA GCACAGATCT TGAAGCAAAA TGCATGAGAG TGTTGGCTTT GAGTTACCAT 1140

CACTTGCCTT CTCACCTAAA ACCGTGTTTT CTGTATTTTG CAATTTTCGC AGAGGATGAA 1200

CGGATTTATG TAAATAAACT TGTTGAGTTA TGGGCCGTAG AGGGGTTTTT GAATGAAGAA 1260

GAGGGAAAAA GCATAGAAGA GGTGGCAGAA ACATGTATAA ACGAACTTGT AGATAGAAGT 1320

CTAATTTCTA TCCACAATGT GAGTTTTGAT GGGGAAACAC AGAGATGTGG AATGCATGAT 1380

GTGACCCGTG AACTCTGTTT GAGGGAAGCT CGAAACATGA ATTTTGTGAA TGTTATCAGA 1440

GGAAAGAGTG ATCAAAATTC ATGTGCACAA TCCATGCAGT GTTCCTTTAA GAGTCGAAGT 1500

CGGATCAGTA TCCATAATGA GGAAGAATTG GTTTGGTGTC GTAACAGCGA GGCTCATTCT 1560

ATCATCACGT TGTGTATATT CAAATGCGTC ACACTGGAAT TGTCTTTCAA GCTAGTAAGA 1620

Fig. 3b (2)

GTACTAGATC TTGGTTTGAC TACATGCCCA ATTTTTCCCA GTGGAGTACT TTCTCTAATT 1680

CATTTGAGAT ACCTATCTTT GCGTTTTAAT CCTCGCTTAC AGCAGTATCG AGGATCGAAA 1740

GAAGCTGTTC CCTCATCAAT AATAGACATT CCTCTATCGA TATCAAGCCT ATGCTATCTG 1800

CAAACTTTTA AACTTTACCA TCCATTTCCC AATTGTTATC CTTTCATATT ACCATCGGAA 1860

ATTTTGACAA TGCCACAATT GAGGAAGCTG TGTATGGGCT GGAATTACTT GCGGAGTCAT 1920

GAGCCTACAG AGAACAGATT GGTTTTGAAA AGTTTGCAAT GCCTCAATGA ATTGAATCCT 1980

CGGTATTGTA CAGGGTCTTT TTTAAGACTA TTTCCCAATT TAAAGAAGTT GGAAGTATTT 2040

GGCGTCAAAG AGGACTTTCG CAATCACAAG GACCTGTATG ATTTTCGCTA CTTATATCAG 2100

CTCGAGAAAT TGGCATTTAG TACTTATTAT TCATCTTCTG CTTGCTTTCT AAAAAACACT 2160

GCACCTTTAG GTTCTACTCC GCAAGATCCT CTGAGGTTTC AGATGGAAAC ATTGCACTTA 2220

GAGACTCATT CCAGGGCAAC TGCACCTCCA ACTGATGTTC CAACTTTCCT CTTACCTCCT 2280

CCGGATTGTT TTCCACAAAA CCTTAAGAGT TTAACTTTTA GCGGAGATTT CTTTTTGGCA 2340

TGGAAGGATT TGAGCATTGT TGGTAAATTA CCCAAACTCG AGGTCCTTCA ACTATCACAC 2400

AATGCCTTCA AAGGCGAGGA GTGGGAAGTA GTTGAGGAAG GGTTTCCTCA CTTGAAGTTC 2460

TTGTTTCTGG ATAGCATATA CATTCGGTAC TGGAGAGCTA GTAGTGATCA CTTTCCATAC 2520

CTTGAACGAC TTTTTCTTAG CGATTGCTTT TATTTGGATT CAATCCCTCG AGATTTTGCA 2580

GATATAACCA CACTAGCTCT TATTGATATA TTTCGCTGCC AACAATCTGT TGGGAATTCC 2640

GCCAAGCAAA TTCAACAGGA CATTCAAGAC AACTATGGAA GCTCTATCGA GGTCCATACT 2700

CGTTATCTTT A***GTAAGACAT CTTCTTCCTT GATTTACAAC AATATTTAAC TCATCATCAT*** 2760

AGTAAACTCG ATAATAATCT GGATAATAGC TTTAGTAAGT CAAATTGCAC CAATTCAACA 2820

AAAGTTCTTG ATGCTGTCAT TGTGATTGAT TCGAATCCTT CCAATATTGT GTAACTTGTT 2880

ATACTTGCAT GTTCATTCTT GATTTTGGGA AGTGTAACAT TTCCATTTTT CATCTTGATT 2940

TTGGGAAGTC GAAATGGAGC ATTTTTGGTA GTGTGA 2976

Fig. 3c (1)

```
CTAGAGATTG GAATGGAGTG ATTCTTAGGG GTTTCTTTTT GAATTAATAT GAGGGTTAGT      60
ATTCAATCTT CAATTCGACA TTTTCTCATA ATTTCTTTAT CTGTTTATTT TTCCTATTCG     120
TAAATCTCTT GGGAAAAATT GGGGTTTTAT CGATTTGGAC TCCTTTTTGA TGAAAAAGGT     180
ATATTTACGA TCTTTATGTT ATGGGTAAAC TGATTTTAAC ATAAAATTAT TGATTCATCG     240
ATTATTTTTA TCATATTAAC CGCGTACAAT TTGGACTTTC CCGGTAAAGT TAAAGTATGA     300
TAAATTGAGA ATTTCAAGGT CGATCTTAGC TCCATTTTTG ATGAAATTTC ATATTTGAAC     360
TTATCTAAGC ATGGGTAAGA TGTTTTTCAA GAAATATTTC ATTTTCGAGT CGGGGTTTTG     420
GATTCGAATA TTTTAGGCTT CTTCAAGAAT GTAGATTTTT GTTTAAATTG AGTTTGTGAA     480
TTGATTTCAA CTCCATTTTC AAATTGGTTT TCACCATTAG CTTCCAAATA CTTTAAGGAT     540
CATTTTACAT CAAAAAATTC CAGATTTGGG TATCGTTTTC CGGTATGAGA CTTTTGGACC     600
GTTTTGCCCC TTTTCCCTAA ATTTCTTGAT TTTGGTGTCA TTGGACTCGA ATTGTGATTG     660
TGAATAATTG TTTGAATAGA TTATCGTGAT CCAGATTATA CTTGGAAAGG AAAGGCTCAA     720
GTCAAGTAAC TTTTGGAGTT CGTTTTAAGG CAAGTGGCTT CCAAACTTTG TAAAACTCTT     780
AGACTACGCA TGACTACTTT CCTAATTATG TTGGGGAGTA ATGGGGGATT GAGGATGGGT     840
TTTATTTGTT GATTGAAATT GTTGTAAATG AAAGATGGGG AATAAAACGA GCTAAATGTG     900
TTATGTGTGA CTTGAATTTG TTTGAATAAG TCATGTGATA ACTGATATTG AGGGATAGAA     960
GAGCATGAGC AGGCTATGAT TGATACAGAC ATTGATGTTG AGGCAGATGA TGTGTAATAC    1020
TATGATGTGG TCGTGATATG GTTGTGATTG AGACATGTGA TGTGTAATAC TATGATGTGG    1080
TCGTGATATG GTTGTGATTG AGACAGGTGA TGTGTAATAC TATGATGTGG TCGTGATATG    1140
GTTGTGATTG AGACAGGTGA TGTGTAATAC TATGATGTGG TCGTGATATG GTTGTGACTG    1200
AGACAGGTGA TGTGTAATAC TATGATGTGG TCGTGATATG GTTGTGATTG AGACAGATGA    1260
TGTGTAATAC GATGATGTGA TCGTGATATG ATTGTGATTG ATTACATGTG CATATTCATT    1320
ATTCATCCCA TGTGTGAACT ATCTGTTGCA TGAGTTCTGA GACACTGATA TGAGGATGGA    1380
TGGATATGAG ACACAGTTGA GACTAGCTCC GGCTAGAGAT GTATGAGATG GACTAGCTCC    1440
GGCTAGCGAT TTGGATGCCG ATGGGATCTG GTTCCGGCGG TGATACATGG TCCATGTGTG    1500
GCCCCCATGG GTTCTGATTT GAGTATTCAA CGCGGACTGA TTACGTCAAC AGATGTGTAT    1560
CGTAGGACAG ACATGTATCA CGACTACATG ACATCATTAT TGCATTTTGC ATCGCATTTG    1620
```

Fig. 3c (2)

CCTTATCTTT GTCTGTGATG TGTGGATTGT ATCGGTTTAC CCTTTTTATG TGGAATTTGA 1680

TCTACTTGCT CTTATTTGTT GATCTGAGGT TGATGAGGAT ATACTGTTGG TTCTGGCTGT 1740

TGAATATGAT CTGTTTAGTA TAGGTTGGTT GGTTTGCTGC TAGATTGAAG TTTCGGTGGT 1800

TCGGTTGGGA TTGAAAGGAG TTGTTTGTAG CTGCTAGTTT TGCTTAGTTT AGAGTTACTT 1860

GCGAGTACCT GTGGTTTTCG GTACTCACCC TTGCTTCTAC ACAATTGTGT AGGTTGACAG 1920

CTCTCTCTCA GATATTTTCT TTAGCAGATT GAGCTTTGAG ACATACTCGA GAGGTAGCGG 1980

TTCATTCCAG ACGTGCCCTT GAGTTATCTT TACTTTCAGT TTTGTTCTAT TCGAGAACTA 2040

TACTCTGAGA CTTGTATATT TTTATTCGAA TTCTGTATTT AGAGGTTTGT ACATGTGACA 2100

ACCAAATTCT GGGTAGTGTT AAGTCTTAAT TAAAGTTTTC TGCTTATTTA TTATCTTTTA 2160

TTCTCGTATT TCTACTTCTC TATCGTTGTG GTTGGGTTAG GCTGACGTGT CTGGTGGGAA 2220

ACGGACATGT GCCATCACAT CCGGATTTGG GGTGTGACAA ATATTTTGTT AGTTATATAC 2280

AAAATTGTAT GTAGTATATG TATATTTTCT GCTTTCATCA CAATTGTATA TAGATATTTG 2340

TATATTTTGT TAGTTATATA CAAAATTGCT TGAAGTATAT GTATATTTTC TGCTTAAATC 2400

ATAATTGTAT ATATATATAT ATATATATAT ATTTCTATAT TTTGTAAGTT ATATACAATA 2460

GTATGAATTA AACAATATAC AAACCTTACA TTATTATATA TACAGTTAGG TTACACCAAA 2520

AATTATCAAA TTAAAGCACA ACTTTTTTAT CGAATCATAT ACAATTCATA TATATAATTG 2580

ACTTAGTAAT TTTATACAAC TACTTACACT TCTACATGGT ATAAGAATTT TGCACAATTA 2640

CTTACATATA TACAATATTA TCAATTAAAC AATATACAAA TCGTATAACT TATATATACA 2700

GTAAAATTAC AACAACAACA ACAAAAATTA TCAAATTAAA GCACACCGTT GTTGTCGAAT 2760

CATATACACT CCATATATAC AAATTGTGTC ATTCAATTTT TCGAACAAAA AATTAGAATT 2820

GAATTGTTAA TATAAAATTT ATCTAATATT GTATAAACAA AATTAAATTA TTGCAAACCA 2880

TTAGAATGAA AAAAACAAAA ATAAACCGTT TTCCAAAATT TCAATTATAT ACTATACAAA 2940

TCAATTGTAT ACTTTCTTGC CGTTCAAAAC ATGAAGTTTC CTTGAAAGAA ACGCTTACCT 3000

AGCGTTGAAT ATACAAGAAT ATTGATTAAT CGTATGCTTC AGTCGTTTGA GGAACCCAGT 3060

TGTTATTGTG TTTCTATTGC TATAGAACTC CTTTTTGGAA AAATATTTGA TTTTGGACGA 3120

TTAGCTTGAA TCATGGGATT ATATAAAATT TTTATTACCG TATTTAGCAC TCATGTATCC 3180

ATTTATTAAA AAAAAATTGT ATAAATTATA TTTTTAAAAG AAAATATACA AAATTAATGC 3240

Fig. 3c (3)

TTCATAGCAA ACTAAACTAT ACCCATTGAA TGTAATTACT AAACTATACC TATAGAGCGT 3300

TATTTCATTA AATACGTTTA TCATATATGA AGTTTTCCCT CAAGAGATCC TACACCTTAT 3360

ATATAGCTTC TCAAATGTGG AAATTCAATC TCACACCCAA CAATCTTTCC CTCAGACTAA 3420

GTTTCATGGC CCAATATCAC AATGATCCAC GAGTCAATTC ATGAGATTCA CTATGTGTGT 3480

CACCCACATC GTCTAAGTAT TTTATGGCAA TCAAGCCCTA CAACTTGCTT CTTCTTTATA 3540

TATATATATA TATATATATA TATATATATA TATATATGTG TGTGTGTGTG TGTGTGTGTG 3600

CGCATCTCTA ATTAATCTCG TAAAGGGATT AAGGGGCCAA TTTCAAAGAA TTAGGCGATT 3660

TTCTTAGTTT TTCGTGTGTG TTAACCCATA GGTATTTTGG TGATATGGTT TTCGGATGAT 3720

TTATTTTGTG CAACTTATAT GGAACCCTTC GTAGGGAGTT AGTCTCACAC TTTTTAGAGT 3780

CCATTTTGGG CATTCAGGGG CTAATTTATA GGAAATAGGT GATCTTCTCA GTTTGTCTGT 3840

ATTAGCCCAT GAATATTTTG GTGATATGTC TTCCGAATAA TTTCTTTGTA AAATCTTTAC 3900

GGGACCCTCC ATAGGGAGTT AGTGGAGCAG TACGTATAGT CTCACAATTT TAGAGTTCAT 3960

TTTGGGCATT TAGGGGCCAA TTTACAGGAT TTAGGCGACT TTCTCAGTGT TTTGTGTGTG 4020

TTAGCCCATT AATAGTTGGT GATATGACTT TCAGACGATT TCTTTGCTAC ACATTTACGG 4080

AACCCTCTGT AGGAAGTCGG GGGAGCAATA CGTACAATCT CACAATTTTA GAGTCCATTT 4140

TAGGCATTTA GGGGCCAATT TAAAGAAATT GGACAATTTT CTCAGTTTTT CGTGTCTGTT 4200

AGCCATTAAT ATATTGGTGA ATATGACCTA CAGATGATTT CTAATCGAAA TCTTTACGAA 4260

ACCCTCAGTA GGGAGTTGGG GGAGCAATAC GTACCGTCTG ACAATTTTTA GAGTCCATTT 4320

TGGGCATTTA AGGGCCAATT TACAGGAATT AGACGATTTT CTTAGTATTT TTTCATGTGT 4380

TAGCCCATAA ATATTTTGTT GATTTGACTT TTAGAGTCTA AACTTCTCAT GTATATTAAG 4440

AGATATTTAT GCTTGGTTAA TTGAATCGAA CTAGGAATAG AGAAATTCCT ACTTGGATCT 4500

TAATATTTCT CTCTCTTTGA TTTGGAAAAT TCTAGGAAGT TGCTTTCAAT GGAATTAAAA 4560

TCATCAATCT CTTGTATGTA AGAAACATAC TTATATTCAT GAATAGATAT GTTTAGGGTC 4620

TAATAATGAA TTATCACAAT TTTTTCTACT TTTTCTTGTC AGAGTCCTGC CTTTTTCTTT 4680

TTCTTTTTTA ACTTTGGTCT CTGCTTTTGT CTACATGATG ATAAGGTTGG TGGACCTAGC 4740

TGGAAATGTG ATGGAAATAG CTAGTAAAAG AAAGAACTTT GCATTTTCTG TTTTCTTAAA 4800

AACTGATAAA TTACATAACT TGTGGCAATT TGTCCATTTT CATACTGAGA GATATTTCTA 4860

Fig. 3c (4)

TTTTTTTTGG ATATATGGCT TATGCTGCTG TTACTTCCCT TATGAGAACC ATACATCAAT 4920

CAATGGAACT TACTGGATGT GATTTGCAAC CGTTTTATGA AAAGCTCAAA TCTTTGAGAG 4980

CTATTCTGGA GAAATCCTGC AATATAATGG GCGATCATGA GGGGTTAACA ATCTTGGAAG 5040

TTGAAATCAT AGAGGTAGCA TACACAACAG AAGATATGGT TGACTCGGAA TCAAGAAATG 5100

TTTTTTTAGC ACGGAATGTG GGGAAAAGAA GCAGGGCTAT GTGGGGGATT TTTTTCGTCT 5160

TGGAACAAGC ACTAGAATGC ATTGATTCCA CCGTGAAACA GTGGATGGCA ACATCGGACA 5220

GCATGAAAGA TCTAAAACCA CAAACTAGCT CACTTGTCAG TTTACCTGAA CATGATGTTG 5280

AGCAGCCCGA GAATATAATG GTTGGCCGTG AAAATGAATT TGAGATGATG CTGGATCAAC 5340

TTGCTAGAGG AGGAAGGGAA CTAGAAGTTG TCTCAATCGT AGGGATGGGA GGCATCGGGA 5400

AAACAACTTT GGCTGCAAAA CTCTATAGTG ATCCTTACAT TATGTCTCGA TTTGATATTC 5460

GTGCAAAAGC AACTGTTTCA CAAGAGTATT GTGTGAGAAA TGTACTCCTA GGCCTTCTTT 5520

CTTTGACAAG TGATGAACCT GATTATCAGC TAGCGGACCA ACTGCAAAAG CATCTGAAAG 5580

GCAGGAGATA CTTGGTAGTC ATTGATGACA TATGGACTAC AGAAGCTTGG GATGATATAA 5640

AACTATGTTT CCCAGACTGC GATAATGGAA GCAGAATACT CCTGACTACT CGGAATGTGG 5700

AAGTGGCTGA ATATGCTAGC TCAGGTAAGC CTCCTCATCA CATGCGCCTC ATGAATTTTG 5760

ACGAAAGTTG GAATTTACTA CACAAAAAGA TCTTTGAAAA AGAAGGTTCT TATTCTCCTG 5820

AATTTGAAAA TATTGGGAAA CAAATTGCAT TAAAATGTGG AGGGTTACCT CTAGCAATTA 5880

CTTTGATTGC TGGACTTCTC TCCAAAATCA GTAAAACATT GGATGAGTGG CAAAATGTTG 5940

CGGAGAATGT ACGTTCGGTG GTAAGCACAG ATCTTGAAGC AAAATGCATG AGAGTGTTGG 6000

CTTTGAGTTA CCATCACTTG CCTTCTCACC TAAAACCGTG TTTTCTGTAT TTTGCAATTT 6060

TCGCAGAGGA TGAACGGATT TATGTAAATA AACTTGTTGA GTTATGGGCC GTAGAGGGGT 6120

TTTTGAATGA AGAAGAGGGA AAAAGCATAG AAGAGGTGGC AGAAACATGT ATAAACGAAC 6180

TTGTAGATAG AAGTCTAATT TCTATCCACA ATGTGAGTTT TGATGGGGAA ACACAGAGAT 6240

GTGGAATGCA TGATGTGACC CGTGAACTCT GTTTGAGGGA AGCTCGAAAC ATGAATTTTG 6300

TGAATGTTAT CAGAGGAAAG AGTGATCAAA ATTCATGTGC ACAATCCATG CAGTGTTCCT 6360

TTAAGAGTCG AAGTCGGATC AGTATCCATA ATGAGGAAGA ATTGGTTTGG TGTCGTAACA 6420

GCGAGGCTCA TTCTATCATC ACGTTGTGTA TATTCAAATG CGTCACACTG GAATTGTCTT 6480

Fig. 3c (5)

TCAAGCTAGT AAGAGTACTA GATCTTGGTT TGACTACATG CCCAATTTTT CCCAGTGGAG 6540

TACTTTCTCT AATTCATTTG AGATACCTAT CTTTGCGTTT TAATCCTCGC TTACAGCAGT 6600

ATCGAGGATC GAAAGAAGCT GTTCCCTCAT CAATAATAGA CATTCCTCTA TCGATATCAA 6660

GCCTATGCTA TCTGCAAACT TTTAAACTTT ACCATCCATT TCCCAATTGT TATCCTTTCA 6720

TATTACCATC GGAAATTTTG ACAATGCCAC AATTGAGGAA GCTGTGTATG GGCTGGAATT 6780

ACTTGCGGAG TCATGAGCCT ACAGAGAACA GATTGGTTTT GAAAAGTTTG CAATGCCTCA 6840

ATGAATTGAA TCCTCGGTAT TGTACAGGGT CTTTTTTAAG ACTATTTCCC AATTTAAAGA 6900

AGTTGGAAGT ATTTGGCGTC AAAGAGGACT TTCGCAATCA CAAGGACCTG TATGATTTTC 6960

GCTACTTATA TCAGCTCGAG AAATTGGCAT TTAGTACTTA TTATTCATCT TCTGCTTGCT 7020

TTCTAAAAAA CACTGCACCT TTAGGTTCTA CTCCGCAAGA TCCTCTGAGG TTTCAGATGG 7080

AAACATTGCA CTTAGAGACT CATTCCAGGG CAACTGCACC TCCAACTGAT GTTCCAACTT 7140

TCCTCTTACC TCCTCCGGAT TGTTTTCCAC AAAACCTTAA GAGTTTAACT TTTAGCGGAG 7200

ATTTCTTTTT GGCATGGAAG GATTTGAGCA TTGTTGGTAA ATTACCCAAA CTCGAGGTCC 7260

TTCAACTATC ACACAATGCC TTCAAAGGCG AGGAGTGGGA AGTAGTTGAG GAAGGGTTTC 7320

CTCACTTGAA GTTCTTGTTT CTGGATAGCA TATACATTCG GTACTGGAGA GCTAGTAGTG 7380

ATCACTTTCC ATACCTTGAA CGACTTTTTC TTAGCGATTG CTTTTATTTG GATTCAATCC 7440

CTCGAGATTT TGCAGATATA ACCACACTAG CTCTTATTGA TATATTTCGC TGCCAACAAT 7500

CTGTTGGGAA TTCCGCCAAG CAAATTCAAC AGGACATTCA AGACAACTAT GGAAGCTCTA 7560

TCGAGGTCCA TACTCGTTAT CTTTA***GTAAG ACATCTTCTT CCTTGATTTA CAACAATATT*** 7620

TAACTCATCA TCATAGTAAA CTCGATAATA ATCTGGATAA TAGCTTTAGT AAGTCAAATT 7680

GCACCAATTC AACAAAAGTT CTTGATGCTG TCATTGTGAT TGATTCGAAT CCTTCCAATA 7740

TTGTGTAACT TGTTATACTT GCATGTTCAT TCTTGATTTT GGGAAGTGTA ACATTTCCAT 7800

TTTTCATCTT GATTTTGGGA AGTCGAAATG GAGCATTTTT GGTAGTG<u>TGA</u> CAACAGATGA 7860

AGATGATGAT GATAGTGTGA CAACAGATGA AGATGAAGAT GAAGACTTTG AGAAAGAAGT 7920

TGCTTCTTGC GGCAATAATG T***GTAAGTTCT TATACCTGCA TGCTCATTCT TGCTATAATG*** 7980

TTCTCTTGTT CCTTAATTAT GGGACATCTA ACATATTATT TTCCATTTTT TGCATCTTTT 8040

TTTTTTCCTG CAGCGTGTAG TTAAGGTGTT CTGAGGACTA GCCAGTTCTC TGAAATAAAT 8100

Fig. 3c (6)

GTCAAATCAG AAGCCAAATG TGTGAGTGTT TGTTTTGTTC GTTTTCATTT TTTCTGCATA 8160

AGGTGGCAGG ATGATTGCAA ATGGCTTGTA ATTTAATTGT ATATGATATT TCGTATAGCC 8220

ATTTGCCAGT GGTTTTTTAG ATACTCCAAA TTTTATGTAC ATACATAATG GTATAGGCCA 8280

GAACAGGCTC CATATATAAC GTGTGTTTCC TTTCTTGGGA GTCCTCAATC TACCTCGCAA 8340

AGGAAGACAG ACGGCTAAAT CAAGAAAGAA ATTTTTTTGA AAATCATGTG GCTAGTTGTT 8400

CAACTTTATA CAAGTTTATG TGCATACTTG TGCATACCCA AAGTTGAATA ACATAAACAT 8460

AAAATGAAGT CAAGTTAAAT GGCACATTTA TGTATTATGC CTTTTGAATT TCATTAATAG 8520

TGAAAATCCT GAATCATATT CAGATTCCAT CACTAATCGT TGAACCATGT TAATTTACTA 8580

TGTATTATCT AATGGATTTT TTTGCTATCT TATTTATAAT TGTTCAAAGT TTTGTTAATT 8640

ATCTTTAGCA TAATATCTGA TTATATTATT TTGATATACT TTCTCTATCC CTAATTACTT 8700

GTCCATTTTT GAATTGGCAC ACCTATTAAG AAAATAATTA TTGAAATAGT GAGTTTACCA 8760

TTTTACCCAT ATTAATTATG AAGTGGATGA ATTAAAAACT CAAGATTTTC AAAAAGTTCT 8820

ATTTTTTTCA AAGTAATAAA CTGACGGTAT AATAGGTAAA AAAAATTATT CTTTCTTGAT 8880

TTGTCAAAAT AAACAAATAA TTAGGAATAA TTAAAAAAAT GGATAAATAA TTAAAAACGG 8940

AGGGAGCAAT ATGTTATCTT TAGCCTAATA ATATCTGATT AATGGCCACC CTAATTGATT 9000

GGATAGGAGA GGATAGACTT GCTTCCAAGT AACCCAAAAT ATAAAAAGTT GACAAAAGGG 9060

TGCTAAATTC GAGACACATG TAGTACTTAT ATAATTCATG TGCGGACTCG TTCTTTTGTA 9120

GTACTCCCTC CGTTCTATTT TATACGTCAC ATTTTTACTT TATACTTTTA TTAAGAAATG 9180

ATGTAGTTTT ATCTTTCTAT TCTTATTTAA TGTTTTCTTA AGTCAATTTT ATAATAAATA 9240

ATGAATATAT TTTCAAGATT AATTAACTAC TCTATCAAGG GTATAATAGG TAAAATATGA 9300

TAATTTATAC ATAAATTTTA TAAAATGACA AGTATTGTGG TCCAACTATT TATAGAAAGA 9360

AATGATATAT AAAATGGGAC GGAGGGCGTT ATAAAGTTGA CTTAAGAAAA CATTAAATAA 9420

GGGTAGAAGG GTAAAATTAC ATTATTTCTT AATGTAAATG TAAAGTAAAA AGGTAACATA 9480

TAAAATGGAA AGGAGGGAGT AGTATTTTCT TGTTTTATTT TACGTGGCAC TCTATTCTCA 9540

TAATCCGTCT TTAAAAATGT CATTTTATTG TAATTGAAAA TAATTTAACT TAAAATTCTC 9600

CATCTACCCT TAATTAATGA AATGATTTAC AATTATATAA ATATATAAAA ATTGTTTTAG 9660

CCTATAATTT TCTAAAATCT TTTTTTTTCT CTTATACATC GTATTAAGTC AAACATAAAT 9720

Fig. 3c (7)

GGAATGGACG GAGTATTTCT TTTATTTTTT TGTCACACCG CCCATATGTT TTCTCCCATC 9780

CCCCAGACCC CCACTATGTA TATTCACTCC TTAGTTGGAT CTGAATTTAG AGTTTAGAAG 9840

CTTCTATAAT AATTTTAGAT TAATATATAA TAATAATAAT AATAATTGAA CTTACAGTAT 9900

TAAATTTATG TGAATCTATA TATATTGTAT TGTAATTTTT TTAATTATAA TTTTAACCAA 9960

ATCAATAAAG CTATTCAGAT GTAAAAGTAT ATATTATGAT TTAACAACAA ATTTCTATAC 10020

GTCTTCCTAA GTTTTGATGC ATAATTTCCT AAAACTCATA AATTTCCAAG TGACTACTTC 10080

CAGTATTACA ATGAGAACTT ATGTTTCGTT ATGGATTTTC TTAGTGAATT AGTTTAATAA 10140

AATCAAAATG AAAAAAAATC ATGTTTTATA ACATAAAATT TTCATTGATT CATGCGAAAA 10200

AAAAACATCT AGTTCTTATA GTGTGAAAAC TATTGAACTT ATGGGATGTA GCTGTATGGA 10260

AGTTCATCAA GTGGTAGCTC CTTGTACGCA ACTAGTGCTA CTTTTTATTG ACTAAAAGTT 10320

ATTTTCTAG 10329